DOCUMENT RESUME

ED 215 008 UD 021 795

TITLE	Evaluation Strategies for Urban Intervention Program. Proceedings from the Workshop on Urban Intervention Programs (Washington, DC, October 23, 1980).
INSTITUTION	Horace Mann Learning Center (ED), Washington, D.C.
PUB DATE	23 Oct 80
NOTE	63p.
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Compensatory Education; Early Childhood Education; Elementary Secondary Education; *Evaluation; Evaluation Methods; *Evaluation Needs; Federal Aid; *Federal Programs; *Government Role; Program Development; *Research Problems; *Urban Education
IDENTIFIERS	Pasadena Unified School District CA; Push for Excellence Program

ABSTRACT

This is a report on the proceedings of the Workshop on Evaluation Strategies for Urban Intervention Programs sponsored by the Education Forum Branch of the U.S. Department of Education and held on October 23, 1980. The report contains the remarks delivered by the workshop speakers as well as the questions and statements of different panelists. Included are discussions on: the Federal role in urban intervention programs and research and development; approaches to evolutionary program development and evaluation of such programs; evaluation methods and the problems of evaluating intervention programs; and elements of urban intervention programs. Various experiences in urban intervention programs are described with emphasis on their implications for evaluation. (MJL)

PROCEEDINGS

EVALUATION STRATEGIES FOR URBAN INTERVENTION PROGRAM

Education Forum Branch
U.S. Department of Education
Washington, D.C.

October 23, 1980

INTRODUCTION
BY BARBARA CHAVEZ KUBAN,
DEPUTY ASSISTANT SECRETARY FOR HUMAN RESOURCES

I want to welcome you to this Workshop on Evaluation Strategies for Urban Intervention Programs. Although my prime responsibilities are in management and in organization, as a parent and a grandparent and as a citizen, I am really interested and concerned about the quality of education in our schools today. I think we all have a responsibility to find ways to help improve educational delivery.

Another concern of mine which I would like to share with you is one that I share with Secretary Hufstedler. I hope it is one of yours or will be one of yours in the future. It is a very special type of school, primarily an urban school, and is called a child care center. There are approximately 7.2 million children of preschool age requiring day care. To me this means that the trend will be toward seeing child care as a responsibility of an extended family, that family being the employer, the community, political and educational institutions.

The educational community, I feel, must take a very active role in developing programs for preschool children as well, primarily children of low income families. I believe that these kids have to deal with our social system, our educational system in the future, and are not prepared for that even before they get into the ABCs. At the Department of Education, we will be attempting to establish a day care center to deal with special educational concerns and needs of this very young and important population.

I mention this to you because as experts and interested parties in the educational field, I would hope that you would include these urban schools in your future discussions for improving educational delivery. I also hope that I can tap some of you as resources as we move on ahead with this very important project.

I would like now to turn this over to Floretta McKenzie, who is the Deputy Assistant Secretary for School Improvement. I want to thank her publicly for continuing to bring us together on these very important issues. If I may, I also would like to thank Grace Watson of my staff at the Horace Mann Learning Center who is committed to bringing us together on these issues. I think between all of us these will be very successful and we can share this information with a broader community.

FEDERAL ROLE IN INTERVENTION PROGRAMS
AND RESEARCH AND DEVELOPMENT
BY FLORETTA McKENZIE
DEPUTY ASSISTANT SECRETARY FOR SCHOOL IMPROVEMENT

I think the applause for Barbara is well deserved, not just because Barbara's budget helps us provide such fora, but because she is definitely committed to the programs of the Department and not to just seeing that people are where they belong in chairs and tables. That is what makes a Department move, those of us who work together for common goals. So, I give a very sincere thanks to you, Barbara, because there is no other way for us to pull together such a wonderful program and such very important individuals to work with us on a very critical topic.

We have had a series of programs or workshops on programs to improve urban schools. But then we thought that we put programs in place and sometimes people expect immediate results. But the programs about which we are talking sometimes do not give us an immediate return. These are programs that are directed toward the prevention of waste, waste of the most valuable resource of this country--human beings. These people, more often than not, live in our urban communities. They are often poor. They are often minorities. We often have low expectations as to what they can do. But as an educator at the local and state level, and as you who are in education, we know that our charge is to help every person develop to his or her fullest potential.

These programs are developed not only at the federal level, but local education agencies have intervention programs, some that we don't even know about. States fund intervention programs about which we do not even know. If we were to put together an inventory of programs that are trying to meet the needs of members of our society who often are alienated, I think we would have a very large volume.

But I see as a part of the role of the Federal government not only the responsibility to provide some of the seed money or venture capital, if you will, to provide opportunities for us to bring new participants into the school scene, that is, business in a different way, to bring in parents in a different way, and also to enable school people at the local and state level to use new strategies, new methodologies to try to meet the needs of a large segment of our population.

There are three programs within the Office of School Improvement that are very visible urban intervention programs: the Baltimore Blueprint; the Cities and Schools Program; and PUSH/EXCEL. These programs are evaluated periodically, and it seems to me sometimes as if we joyously await information about the failures, or what it is that the programs did not do. We do have people here who are in the business of helping us look at these programs. In talking to the people who are in the business of helping us look at the programs, I think they are just as committed that we find out those things that are working well. But unfortunately others only highlight those things that are not working well.

One time I was working in a school system and so many things were going wrong. I have a partner here who was with me at that time. It got to the point where we were trying new things almost monthly. When you know that you are not meeting the needs of a population, you do have to try different things. So, the Federal Government must be into risk taking. We are not only dealing with those things about which we know a lot, but we are in the business of expanding the body of knowledge, developing the state of the art. That is what you are about if you are into programs that deal with improving urban education. We just do not have all the answers, and we must not be frightened away from finding the answers because the research or the evaluation tells us at this point that maybe this is not working. We cannot always also be concerned about how many are we impacting, what is the cost. Once you come to the solutions, then you can start weeding out some of the cost factors.

We as a Federal Government are into small intervention projects on the urban level and we are into some large ones. The things we do in Title I, the things we do in ESEA, all are directed at improving the quality of education and thus the quality of life for many people in our urban centers.

FEDERAL ROLE IN INTERVENTION PROGRAMS AND RESEARCH AND DEVELOPMENT BY MICHAEL TIMPANE, DIRECTOR, NATIONAL INSTITUTE OF EDUCATION

Oh, my God, you might say, researchers are going to tell us how to develop programs again. That was tried once, or more. There is some feeling that some of the disappointments of the 1960s might have been caused by a little over-arching ambition or over-confidence on the part of researchers who thought they knew exactly what we should do.

I guess one of the things I want to say today is that we have not been inactive during these years. Believe it or not, even researchers can learn something from experience about their own methods as well as about the development of programs. I think this is an important underlying message in what Norm Gold and others of my staff who are here today will be saying. We have come a long way in 15 years in terms of understanding how to identify and interpret these complicated phenomena which make up the urba. education scene.

We also have a new opportunity to demonstrate that new expertise. That new opportunity is the Education Department and the Office of Educational Research and Improvement in which we sit. We have thought from the outset that OERI offered an opportunity for us to make the fruits of our research more easily accessible through the organization of the federal education activities to those who were developing the most risky but perhaps the most exciting new programs. We felt we would be in

closer relationship with people like Floretta McKenzie and with the others in the programs of the Office of School Improvement and, indeed, the other demonstration programs of the Department. That, indeed, is turning out to be the case. It pleases us greatly, because we think we have information to contribute, and the distance across which that communication needs to take place has shrunk greatly. This is very good.

Research, or at least the research NIE does, has a couple of things to contribute to the kinds of programs whose assessments we are going to look at this morning. The first is that year in and year out at the Institute we are conducting research based on trying to improve the state of the art and the delivery of education services themselves. So, we are doing fundamental research on how kids read and how they learn math. We are doing a great deal of research on how tests are used. We are doing lots of new research on the acquisition of language. We have spent a great deal of time over the past several years trying to figure out exactly how schools as organizations work. These are all more general themes of research which play in over the years to our understanding of how the particular kinds of projects about which we are going to be talking today can work or might work.

The second brand of activity that we carry out is evaluation itself, evaluation of ongoing projects. Norm Gold will be speaking to you in some detail about the methods we have developed to conduct such evaluations, because we are very sensitive to the particular needs of innovative urban education projects to an evaluation which has, to start with, carefully identified the objectives of the program and just what would or would not constitute criteria for success. No research or evaluation effort can come charging in to the dynamic and living scene of an urban intervention and try to figure out exactly what's going on without long and careful discussion and understanding of what it is that is intended in that intervention. So we have developed a fair handful of evaluative efforts of this sort.

About five years ago, we began to evaluate the career intern program of the OIC, which, at that time had been set up in Philadelphia. Howard Resnik is here today to tell you that we have since taken that intervention to four other sites at the request of the Labor Department and are now doing one of those great rarities in the history of research, an honest to goodness replication and evaluation. We have taken a successful intervention in one site and have installed it in four others, and evaluated it to see whether or not its success is attributed to the site in which it first succeeded or to the program itself.

We have been evaluating two of the exciting programs that Ms. McKenzie mentioned, PUSH/EXCEL and the Cities and

Schools program. I am sure Norman will speak from personal knowledge in great detail about those.

We developed the Experience-Based Career Education Program a few years ago, about which many of you know, to try to develop new models for vocational training. We concentrated especially on the urban scene.

Most lately, we have participated a good deal in the development of the youth initiative. As you know, it has passed the House of Representatives and will be pending on the Senate's calendar when it returns after election day. This is an attempt, once again, to develop a new intervention effort, aimed at the secondary schools in our cities. This is a desperately needed new focus, it seems to me. I think we have been able to contribute a good deal to the design of that program in a way which stresses the literacy and employability skills that it seems to us the kids in those schools need more than anything else. It stresses the need for the whole school to be involved in changing its programs fundamentally, not in adding on another program off in the corner somewhere. We hope to begin to impart those skills to those young people. That is a very exciting prospective intervention that we hope to continue to be involved in.

In the last few months we have begun to forge a partnership which excites us very much with the Follow Through program in which we may be able to contribute to that program's efforts to develop new or modified models of intervention for the programs for the primary grades, which Follow Through has long fostered in our cities, so that we may again contribute some of the basic insights we have gained on effective instructional strategies and effective organizational strategies in our schools to make these interventions sail.

So, being here today we bring to you, we hope, experience from half a dozen major effor. to see how the tools of research and evaluation can bear upon the success of these urban interventional programs. We want to tell you that what we have learned has been learned continuously over the past 15 years, and what we have to say today, as I am sure Eleanor Farrar's remarks will reflect, is a different story than formerly was told.

APPROACH TO EVOLUTIONARY PROGRAM DEVELOPMENT
BY ELEANOR FARRAR,
VICE-PRESIDENT AND SENIOR RESEARCH ASSOCIATE,
THE HURON INSTITUTE.

Evaluation used to be a modest cottage industry in American education, but now it has become a big business. Twenty years ago, a meeting on urban school reform would have been con-

cerned with curriculum revision or training teachers and ten years ago, a meeting on school reform would have been concerned with reforming the organizational climate or with improving interpersonal relationships among students and teachers.

But today we are here to talk about evaluation. School researchers and practitioners alike seem to be almost exclusively concerned with program evaluation and assessment, and issues that involve testing. Program development seems to have gone out of style, particularly directed development efforts. But I think it makes sense to pause for a moment to think about just what it is that we are evaluating, what it is that we propose to apply our evaluation methods to.

The school reforms of our time are far more comprehensive and complicated than they were a decade or two ago. The relatively straight-forward curriculum reforms of the early 1960s involving just students and teachers-- curriculum reforms which, it turned out, when Bob Stake took a closer look at them, were not so straight-forward after all--seem simple in their conception and design by comparison. The new generation of school reforms involves not only students and teachers, but parents, members of the community, occasional social service agencies, and increasingly employers as well. The new wave of school and work programs is a good example. Cities and Schools is another example, as is PUSH/EXCEL and Experience-Based Career Education. These programs tend to blur the boundaries between schooling and community. They are enormously complex social interventions and they involve many actors.

So, at the same time that the state of the art of evaluation has become more complex and the methodologies of evaluation have become controversial in their application, the programs have become more complex. So have our understandings about these programs and the way that they are developed and implemented. Nearly gone are the days when people think about evaluation as simply a three-staged process, in which you pretest the kids, you apply the treatment under some kind of conditions that are as nearly controlled as you possibly can get them to be, and then post-test to see if the intended treatment had its effect. By the same token, simple conceptions of program development have been swept away by much more elaborate formulations about how the process works.

I'm going to talk about both the old and the more recent ideas about program development and implementation and our notions about how this process occurs because I think it has important implications for program evaluation for what purposes these evaluations realistically can serve in providing information about program progress or success.

In brief, I would like to suggest that programs are evolutionary in both their development and in their implementation and that, for all practical purposes, there are really few distinctions between program development and implementation. Programs developed in R&D laboratories or in demonstration schools are further developed once they get out into local districts and schools. The process is continuous. The more familiar view, the inherited view, is that one stage follows the other, that the program is developed in a laboratory, it goes through some kind of development stage and then is implemented, ideally in a more or less high fidelity fashion in local schools. But what I am proposing is that there is no fixed entity to be implemented by schools; rather, the schools develop variations on a set of policy ideas or improvisations on a program design which is provided by the Federal Government or provided by program developers they supported.

I call it a theme because at base it is a set of policy ideas or intentions about improving schools, intentions that have been shaped into operating programs. The themes may have form and structure and set procedures, that is, they may have a blueprint, such as the ExperienceBased Career Education model or the Career Intern model, or it may be just a loose set of ideas about how to improve education, something that comes with guidelines and regulations to steer local development, such as the Youth Act. It might be very specific or it might be relatively flexible and unformed. But, in either case, local implementation gives a new or a different shape to the program or policy intentions and the program evolves.

Development of implementation is an evolutionary process with a beginning, but a very uncertain end. But if programs are thought of as evolutionary, how can they be evaluated? What are the criteria one uses to determine success or failure? At what point in time do we decide to evaluate and for what purposes?

The answers to these questions seemed relatively straight-forward when people held a simpler view of program implementation. The old ideas informed most program development efforts during the 1960s and right through most of the past decade. It was thought to work in the following manner. The Federal government--usually it was the Federal government--established a set of policy ideas for reforming American schools. These were usually a bunch of ideas that had been floating around in the social science literature at the time, and they were ideas picked up by the Federal government. The Federal government funded developers to turn these ideas into a bunch of operating programs that could be implemented locally. These programs were developed in research and development laboratories or in demonstration schools or lighthouse districts and they resulted in a set of plans or blueprints, and they were accom-

panied by curriculum guides and instructional materials and detailed procedures for local practice.

Sometimes the programs were accompanied by training, sometimes by technical assistance from the developers and sometimes they were not accompanied by these things. But in most instances the programs had undergone extensive evaluations and assessments to determine that indeed they worked, and indeed they were ready for broad-scale implementation. The programs were certified.

Once the programs were determined to be ready, it was expected that schools and school districts would implement it pretty much as was planned, using the materials and procedures and embracing the same goals for evaluation. The ideal was a high fidelity replication of the program. Local schools might make a few modifications and changes in the program or in the blueprint to accommodate local circumstances, but they would not tamper with the innovation in major ways. It was proven to be effective only if it were implemented as the developers had intended.

This was a handy way of thinking about program development. In particular, the notion that programs were more or less fully developed before they reached local schools was consistent with the methods of systematic evaluation studies. This view of programs was sympathetic to the methods of science. There were formal blueprints and procedures to use in assessing the effectiveness of implementation. The programs had specific goals, often linked to outcome measures, and the treatment was directed toward the goals. Students were pre-tested and post-tested and the lack of gains presumed that there had been some kind of breakdown in implementation. People were not doing what they were supposed to do. Failure to achieve the desired results was caused by failure to carry out the program as intended.

Arguments sometimes ensued about the appropriate design for the evaluation or the appropriate goals for the program. But one thing seemed clear, and that was the program, as it was developed, represented the best way to implement federal policy in a way that had been certified by program evaluations. Thus the program design was completed once it was out to local schools and in that form it was stable. If it were stable, it could be evaluated using fairly rigorous methods.

These ideas about program evaluation began to give way in the 1970s at a time when evaluations of most studies began to suggest that they were not succeeding and when implementation studies revealed that local districts were making wholesale modifications and changes in programs. Ideas began to shift at a time when increasing attention was being given to the role of local actors in implementation. Research on behavior in organizations and several federal studies on implementation suggested that it was more sensible and more realistic to pay attention to

the fact that local people have particular needs and preferences of their own to which the program must be accommodated, preferences which were not always consistent with those of the developers or those of the Federal Government.

In addition, people began to talk about local contexts. The program was being introduced to a very particular local context that varied from district to district, and it only made sense that there would have to be changes to accommodate to local context if the program was to thrive, never mind survive. Furthermore, the local context constantly was changing as a result of personnel, political, and economic changes, and experience had an impact on implementation as well. As local people began to adopt and implement the programs, they tried to improve on what they did. They changed it.

The upshot of all this was wide modification of program blueprints and increasing reports that replication was not working out. Components were massively revised or sometimes abandoned altogether; curriculum materials sometimes were used, were used by some staff members and not by others, or were revised to meet the needs of local students. In effect, local staff redeveloped the program to meet local circumstances. Sometimes these redevelopment efforts were well planned and carefully planned; but, more often, staff stumbled upon or invented new ways of doing something and they did those things in a somewhat serendipitous way. In many ways, though, those were developments. The program was continuing to be developed, and, as the programs continued, they were implemented.

It seems a more accurate description of reality to consider program development and implementation as part of the same process, rather than two separate stages. It is an evolutionary process as program staff choose to do what they think is best for their particular setting. As they make decisions about the program and as they act on those decisions, the program evolves.

This way of thinking about program development raises two big questions. One is the following. If programs are evolutionary and implementation inevitably leads to change, what does this mean for evaluation? What criteria do we use in judging success or failure? The program's survival is one criterion. That was one that biological evolutionists used. But it does not tell us enough. Survival is important, but survival in what form -- an improtant question, particularly when the programs are being supported with federal money year after year.

Another question that this view of program development raises concerns the investment of large amounts of federal money into directed development efforts. If thousands or millions of dollars are spent on curriculum development or program R&D to

produce a product for local schools, is it a wise investment, given the wholesale revision and modification of the products that ensue?

I'll tackle the second question first because I think it is one of the more important questions for education policy R&D these days. My answer is yes, but investment only up to a point. The directed development of programs for school improvement is important because the programs provide local staff with a program prototype, with a conception of the program that they use as a central frame of reference for their own implementation activities. It provides guidance. It's a blueprint to be tried in parts in varying degrees or in its entirety. It enables start-up to occur more quickly and by people who do not necessarily have a clear idea of what form the innovation might take. Many of the people who are involved in these programs are not educators and they are not familiar with education programs.

Education program prototypes introduce new ideas, new materials, and new methods to schools. Though perhaps only some are used, they encourage new organizational arrangements by describing them. District staff may accept or reject the blueprint, but regardless, it stimulates a whole bunch of ideas about what will or what will not work in that particular setting. Developed programs provide local districts with a theme, a theme that will produce many improvisations. Sometimes the improvisions are more appealing than the theme. That certainly has been the case in musical or literary history. But the point is that these improvisations represent a creative interaction between a central idea and those of local school people striving to improve on the theme or to elaborate it to fit a set of local preferences, local style or way of going about things, or local context.

The view of program development and implementation as evolution and improvisation increases the need of local staff planning time, staff training, and technical assistance. Planning time is particularly important with this new generation of programs that involve many diverse interest groups convened together to work on issues that are unfamiliar to many and in ways that are new and untried. They need to get together to work out new organizational arrangements, to plan sensible division of labor, to decide on areas of shared responsibility, how things are to be carried out, and so forth. It is a long list with which we all are familiar. The more complicated the programs or the initiatives and the more diverse the groups that they involve, the more planning time is required before local groups start to implement the program with students.

Technical assistance and training also are important in the implementation of complex school intervention programs. Both ordinarily focus on helping with the installation of prespecified programs. But broader conceptions of technical assis-

tance may be necessary when a great deal of coordination between organizations and groups is required.

Technical assistance is trouble-shooting, hand-holding, counselling, expert advice -- all of those good things that are desirable in any kind of complex work, those things that we all wish we had gotten and sometimes do on our own. It is particularly necessary when many participants are unfamiliar with the new roles into which they have been cast and they have little experience with these programs or others like them. It is particularly important in weakly specified, broad-scale social innovations, when there is no programmatic frame of reference or blueprint, with the usual accoutrements to guide implementation.

There are lots of good reasons for providing local schools with implementation assistance during the early days of a program's life cycle. But, suffice it to say that the evolutionary view of program implementation does not obviate the need for assistance. If anything, it enhances the need, but not for assistance of the usual sort. People need not only instruction on how to replicate something that somebody else has devised, but they also need help in devising improvisations of their own. They will improvise in any event; but in many cases, the more ideas they have, the richer the improvisation might be.

This brings me back full circle to the question of what do we do about evaluation when we conceive of programs as evolutionary rather than as pre-specified for uniform implementation. If the blueprint is not what is being implemented, what serves as the frame of reference for making assessments about effectiveness? If program goals shift as local staff find that there are some things they can do better than others, or as they find that some local students need some things more than others, what do evaluators look for? Pre-test data may become obsolete as program goals change. Time frames for evaluations may make it seem as if not much has been accomplished, and all the while, theological arguments wage about the proper methods to use in studying programs of this sort.

My answer is that there is no simple answer. If one approaches the matter of program development explicitly as an exercise in diversity and change, one cannot expect stable points of reference. Old-fashioned evolutionists solved the problem by asserting the survival of the fittest or some other doctrine of progress, but I propose no such thing. One possible, but only partial, solution is to admit that there are varieties of success and that no single criterion can comprehend them. Another possible, but quite uncertain, solution is to try to frame processual criteria for success, using such notions as problem solving to judge whether good things are happening. But another is to admit that no single intelligence can comprehend

all the possible ways that programs can be successful. Many different criteria representing interests of many people would tell diverse and divergent stories about program success. But if we can envision these and other solutions to the new problems of evaluating programs, it would be rash to pretend that we could pick the best one.

DR. JAMES COMER, Yale University: I did not understand your last term, processual, I believe. What do you mean by that?

DR. FARRAR: If one possibly could figure out some ways of assessing process, a process that we could use, as a way for carrying out some kind of program or making decisions about it and determine the process to be effective or not.

DR COMER: And then looking at stages within the process?

DR. FARRAR: Well, I was not so much thinking of stage theory as much as I was thinking about what are constructive and unconstructive ways to go about problem solving or decision making.

RICHARD ZUSMAN: I'm concerned with your last statement about the process. If you have limited resources and you are trying to make a decision of where to put those resources, where to put those dollars, how do you make that decision if you just have process information about the implication of a program?

DR. FARRAR: I think the decision would rest, as so many decisions often do, regardless of whether there is hard data or not, on what we might call clinical judgment. You have a notion about a process that seems to be working. I mean, God knows just exactly why it seems to be working, but you have a sense that there is a process, a healthy process or constructive process of decision-making that's going on and you decide to go with it.

MR ZUSMAN: I guess I still feel uncomfortable about that. You're saying they have three healthy processes going on.

DR. FARRAR: That's right. Hopefully there would be more. If you only could fund one, then I would be glad not to have your job. But you're quite right. There could be a whole bunch of them.

DR. COMER: What are the criteria for health? It seems to me that there has to be some stage consideration and that at some point you would determine that although it looked healthy, perhaps it wasn't and didn't get too far.

How would you determine whether or not a program were healthy and what criteria would you use, the process?

DR. FARRAR: I think what I would do, first of all, is to establish a reasonably long planning time, expecting that during that planning time or pre-time, call it whatever you will, a time when people are trying to get themselves together, there might be a lot of interactions, exchanges, or behavior that we would call nonfunctional if we saw it happening three or four years later. But at least in the early stages we let people mess around while getting to know each other and decide what they want to do. At this point we would have some notions of what a constructive decision-making process is like for that particular climate or that particular context. I'm sure it would vary. I'm sure that what would be a good decision-making process in Buffalo would be different from one in Detroit which would be different, in turn, from one in Salt Lake City. It would not be anytning that was uniform. The criteria would seem to be whether it was working for local people, whether they thought it to be something that was constructive and forward moving.

MARJORIE HOACHLANDER, Institute for Museum Services: As a researcher, I must tell you that you have stimulated my mind early in the morning into what I will call a theoretical taffy pull. I do agree with you about the historical change in the application of established programs in various localities. It's almost like a fish swimming this way, saying hi to a fish swimming that way, because the other fish goes into Michael Timpane's world of making NIE empirical evidence to a considerable degree to substantiate not only the funding, but the replication of more things in more places.

That is where my taffy pull rests. If you accept your theory or your desire to do this in this much more flexible, diverse way, how then do you maintain some evidence of empirical documentation? How do you get to that point where somebody else, who has not tried anything before, begins to think in terms of empirical evidence because that is the way some of us are being alerted in other areas of our thinking? How do you come to that kind of mature compromise and design?

DR. FARRAR: I think as a starter the more people who talk about it, the more it will help, people who are solely riveted on empirical evidence, to perhaps pay attention to other kinds of evidence. But I think that it's also useful in evaluating a program to do more than look at empirical evidence, and by that I mean quantitative evidence.

MS. HOACHLANDER: I'm not being solely dependent on that. I'm saying how do you combine the two into a recipe that cooks.

MR. FARRAR: I think that different methods and different points of view can be brought to bear on the study of any single program. If you take a single program, like Cities and Schools, one could try to evaluate it quantitatively, empirically.

MS. HOACHLANDER: That's not synonymous.

MR. FARRAR: No, no. Quantitatively. One can apply qualitative methods to it. People who do both of those things can be evaluation researchers, social scientists. Evaluations also could be done perhaps by social scientists from the perspective of other people who are involved in the program. They could be done from the perspective of teachers who are working in the program, kids who are in the program, parents, so on and so forth. Then what you have is an array of findings.

Now, that does not lend itself to a neat bottom line. If somebody wants to know is it succeeding or is it not, then I think the answer would be well, if you talk to so and so, it is this, and if you talk to such and such, it is that. It is messy, there is no doubt about it. But these programs are messy.

MS. McKENZIE: They sure are not neat for us at the federal level, and I see the members of the national PUSH staff here, and others, and they know that they are not neat either, since they are working with so many different communities.

EFFECTIVE METHODS OF EVALUATING URBAN INTERVENTION PROGRAMS
BY ASA HILLIARD,
CALLAWAY PROFESSOR OF URBAN EDUCATION,
DEPARTMENT OF EDUCATIONAL FOUNDATIONS, GEORGIA STATE UNIVERSITY.

In this short period of time that I have, I have probably chosen to cover too many things, which means that in some places I will be necessarily broad and in others maybe overly specific. But I chose to err on the side of trying to do too much rather than too little, so please forgive me if I am not as complete with anecdotes as I could be.

Our educational system has yet to become distinguished in the sense that achievement level for most public school students can equal that of their counterparts in several other nations. As a nation, we have not yet been able to develop a mass public school education program of the highest quality. Where we do have high quality, we have it for the few and for the privileged.

A Phi Delta Kappa article in the current issue reports that 98 percent of the students in the Soviet Union complete high school, a ten year program, rather than twelve, which includes calculus and a foreign language for the vast majority.

At the same time, the article reports that our own schools produce graduates of only 75 percent of the students. More interesing is a report of a NIE survey which revealed that over half of the nation's schools do not require one year of mathematics for graduation. Japan and Germany also outperform us in public education. Yet, it is our nation more than any other which spends freely of its resources for schools of every conceivable type.

If our public school system as a whole, has been undistinguished, our evaluation of these schools formally has been even less distinguished. Little has been offered to educators from evaluators which can be shown to have improved pedagogy. There has been no dearth of literature in evaluation. No other nation has been so preoccupied as ours in this area. And yet, if anything, we have created mysteries where none existed before. Then, the more we explore the mysteries of pedagogy, the deeper the mysteries have become.

Few urban school systems, if any, are thought of with envy and pride. Yet, the bulk of educational research has been conducted on populations in urban areas. Researchers have described the magnitude of failure in urban education, but have failed in general to describe its dynamics. If we appeal to evaluation data at all as educators, it tends to be as source material for rationalization, for curiosity, sometimes for punitive leverage, but certainly seldom for pedagogical guidance.

Look at the popular studies. Schools do not make a great deal of difference in children's achievement -- Coleman, Jensen. Headstart programs can't change things for poverty children since poverty children can't be changed -- Jensen and Westinghouse. As one looks at older literature in particular, one should not be surprised that the scholars reflected the socio-political thought of their eras and developed expectations much like some of the teachers of low income children. However, one should be surprised that natural and parsimonious questions were asked so late.

I cannot help but be impressed at the profound simplicity of Ron Edmonds' approach to the study of urban schools -- simply look at the ones which do well and find out why. Similarly, the Beginning Teacher Evaluation Study and others like it provided a model for inquiry. These simply have not been applied fully to urban school research. The work of Ray Rist, Rosenthal and Jacobson and many others have shown empirically that life in classrooms is anything but standard for children.

Why did so many evaluators ignore, overlook or miss this important point? The point still is being missed in the main, inspite of such important work as cited above. I believe there are many things about the way the work of evalaution is

conducted that make for major impediments to the work. Yet, a parsimonious look at these things should illuminate basic problems all at once. To take this tack would be to guarantee the improvement of the evaluation enterprise overnight, and therefore would improve urban education evaluation as well. I hope that I can describe a few of these things sufficiently for us to see them in somewhat more bold relief.

Number one is street language in academic evaluation. Every first year graduate student knows that terms must be defined if research is to proceed. Presumably, evaluators who sometimes teach them know this, too. Yet, while the evaluation designs may be constructed with unbelievable rigor at certain points, other key things in the design are frequently treated very shabbily. My mother always told me that when you took a bath and put on clean clothes, you had to use clean underwear as well. The same principle ought to apply in evaluation. For example, what does school success mean in an evaluation design? What is a program? What is race? These terms seldom have either precision or common definition from one evaluation project to another.

The consequences of this, however, are tremendous. Let's take the word "program" for example. Headstart is a program. Teacher Corps is a program. Follow-Through is a program. Title I is a program. But really, these are simply names of categories of fiscal support. They cannot be described uniquely in operational pedagogical terms for some of the reasons that we heard earilier because a variety of things may be done, not necessarily bad, under any one of those headings. So there is no real problem with that unless evaluators want to say something valid about "the program."

Most research on urban education, by its failure to articulate the unique pedagogy of urban intervention, allows for the development of erroneous, implicit assumptions that urban intervention programs offer unique or better treatment than some norm or group with which it may be compared. This was a problem for example, with Jensen's research on the boosting of IQs. At no point in his analysis did he deal empirically with the nature of the treatment which children whose IQs were not boosted had received. Rather, he looked only at child factors for an explanation of his findings. He does the same thing in his new book, Test Bias.

Evaluators pro and con in urban education program evaluation have been consistently sloppy in overlooking or in failing to describe the real "treatment" which the subjects actually get. When evaluators use the language of the street and allow a single street label to represent a wide variety of operational definitions, then reliability sufficient for scientific measurement cannot be obtained.

Second is content-valid testing instruments. Andy Porter and his colleagues should have startled the evaluation world with the presentation of his very rare empirical data on the content validity of standardized tests of achievement. Porter and his colleagues were able to demonstrate that there was almost an average of a 50 percent mismatch between the topics that were covered in popular fourth grade arithmetic test books and popular standardized tests that were used to evaluate the children's progress. One wonders if it is this bad in the well-defined and somewhat sequential arithmetic content area, what we would find in the language arts or social studies areas.

But I wonder even more why there are so few empirical studies of content validity against which to compare the Porter findings. It seems too patently obvious that th. content validity of program evaluation instruments must be es-ablished, not assumed, or accepted on faith, or based on expert opinion. It is the failure to do this simple thing that has resulted in the use of such tests as the Peabody Picture Vocabulary test and the Preschool Inventory Scores as criteria measures for Headstart and other educational interventions where there was no empirical determination that any program under the Headstart label had those criteria as program treatment objectives.

The same may be said of the use of SAT scores for assessing "school success." No matter how laudable the goal of student achievement on the SAT may be, the content mismatch with school curricula will yield invalid results. I am certain it can be demonstrated empirically that the mismatch exists. In short, valid criterion measures are an indispensable prerequisite to evaluation effort.

Third is testing for status or for change. Again, many of these things are things that we say we know, but when we operate, we do not act as if we know. Urban intervention programs require tests with content validity for the intervention, but which are sufficiently sensitive to provide discrimination among levels of achievement along all parts of a total range. Moreover, the sensitivity must be sufficient to detect changes in the performance over short and long periods of time. After all, the rate and quality of change is a cheap item of interest in program evaluation.

Next is outcomes or process. Urban intervention program evaluations frequently are little more than a formative determination that some promised process is in operation. There is little accountability to such a scheme, either for the program or for its evaluator. The effectiveness of urban intervention programs must focus on learner outcomes in the final analysis with process evaluation as a part of that analysis.

A friend of mine who was funded for several years to conduct an innovative teacher education program for urban areas had abundant data on how well the trainees in the program mastered the program competencies, but no data at all ever were collected to determine if the mastery of those competencies by the teachers and the application of the competencies with their students resulted changed student outcomes. This person never attempted to initiate such data collection nor did any donor ever require such.

Next is levels of aggregation. Frequently, the things which could be learned from even the meager data which are collected on urban intervention programs are collected in such a way as to be totally confounded.

For example, it is important to associate the learner outcome data with the specific treatment which is or which could be responsible for the learner outcomes. Yet, data on student achievement are collected frequently and aggregated at least one level beyond that where the treatment actually occurred. So learners experience classroom activities or small group activities within a classroom within one classroom for the most part. Every learner experiences it within one classroom not in the district. They do not experience all the classrooms in the school. Yet the achievement data may be aggregated at the school or even at the district level, obscuring essential information.

In fact, it is this very practice of reporting average performances by school or by district which was contributed to the widespread belief that children who live in urban areas or in culturally homogeneous neighborhoods are uneducable. Disaggregating the data would reveal in many, many cases that the story is not uniform for all classrooms. Many teachers have no trouble at all teaching the "unteachables." Many schools as a whole also have performance records which may be lost in district-wide aggregation, as Ron Edmonds' work has shown.

At a national level, it also is true that performance may vary widely from one project site to the next. I can give another example of a different type. A national agency has collected statistics on the proportion of black children who are assigned to classes for the mentally retarded. They also have figures for males and for females in their breakdowns. Yet, they did not set up their data system to yield the most single important fact of all, which any clinical inspection of EMR classes would suggest. This fact is how many or what proportion of the children in EMR classes are black males. The particular aggregation which was selected obscured the main event. For example, the radical disproportion of black males in EMR classes undercuts the Jensens thesis of the racial basis of IQ test score performance, or at least for EMR placement.

The problem here and with other things, obscured through aggregation, is frequently there is a type of political sensitivity which prevents appropriate aggregation from being done. There is often a fear that the differential performance among classrooms should not be singled out for review, sometimes on the assumption that the bad material, meaning the children, is not randomly distributed, and therefore a good teacher may unfairly be made to look bad, simply because he or she has bad material. Also, in a case such as the case of the black male disproportion, the differences are so blatant that the interpretation may force evaluators and others into sensitive territory for explanations. The aggregation level must match the treatment level at least. Then you can do whatever else you want.

Next I will discuss mythical sequences and mythical components. We must take note of the widespread tendency to think of student learning as following sequences or as being composed of components which do not come from empirical observations of learners. I'm really happy to hear what Mike was saying about the direction of NIE research, but that has not yet caught up with evaluation. These sequences and components appear to be more a product of the logical extrapolation from assumptions about learners than they are from empirical observations. When evaluation activity is expected to give data for diagnosis, we must question the validity of certain popularly accepted sequences and components.

For example, "tests of reading" frequently include not only a test of comprehension, but a test of so-called reading skills as well. The reading skills are presumed, not observed, to be prerequisite to the ability to read or to comprehend. Such skills as word attack or detection of beginning and ending sounds may be a part of a reading test score; and yet, while the particular skills on such tests may well represent good speculation on how learners learn to read, the need is for knowledge of sequences which come from empirical observation.

As we observe how children learn to read or learn to do mathematical operations, it is clear that there are great varieties of sequences among successful learners. One is struck, for example, by the qualitative distinction between Piagetian descriptions of sequences in learning or learning operations and the sequences which are suggested by the components of standardized tests of reading. If Piagetian observation techniques were applied more extensively to reading acquisition, I feel that a very different and more valid content for reading tests would be suggested than that which we now have.

The same may be said about presumed components of behavior, such as Bloom's taxonomy and its cognitive, affective and psychomotor components. This popular taxonomy does not derive from observations of behavior so much as from logical

extrapolations from assumptions about behavior. Again, when one uses Piaget's designations, there are actual behavioral protocols, movies, videotapes, which will illustrate the operation which Piaget has described. I am aware of no such observational backup for Bloom and [Kraftwall] schemes.

Thus the cognitive area, in particular, looks more like an outline for a course in logic than a description of thinking. The elements in the outline may be goals or objectives as they were originally intended for teaching, as the title indicates, but they are not components of thought or sequences of development or even necessarily hierarchical in the sense that some professionals have come to think of them.

The heart of the matter is this. We need valid measures of achievement if evaluation is to be vaild. If evaluation also is to lead to diagnosis, we need valid measures of thinking which would allow impediments to thought to be identified. However, evaluators should not allow these two things to be confounded.

Psychology-bound evaluation. By accident of history and by habit, psychology has come to dominate the field of evaluation in urban education. Psychologists were the first behavioral scientists to be asked to develop an applied approach to the solution of educational problems. Psychologists have claimed the domain of mental testing as their turf. Psychologists have extended and refined statistical methodologies and paper and pencil test construction.

However, a complete understanding of school and school intervention processes cannot be gained by reliances upon the tools of the psychologist alone. Over-reliance on limited psychological approaches to urban educational evalution is responsible for the poor understanding that we have of schools and how they work. Psychological tests of achievement may help to tell us what students know; they do not tell us why. In fact, they do not even tell us all the what.

For example, we will learn virtually nothing about socialization of communication patterns in schools from psychologists. Anthropological methodology, such as ethnography and participant observation are but two examples of powerful data gathering approaches which have illuminated school practices.

If urban education in particular is to be understood, both the what and the why, the psychologi l monopoly on school evaluation must be dismantled. We must have a multidisciplinary approach to the urban educational assessment. Anthropologists, sociologists, cultural linguists and others must join psychologists for the collaborative design of educational evaluation in

urban areas. I am aware that we are ill-prepared for such collaboration at present. We do not have the disposition; we do not have the models; and, above all, we do not have the political climate among professionals. Yet we desperately need such collaboration.

Low expectations for children in urban schools are associated with teacher behavior which can be observed. The differential quality of academic preparation of teachers and different quality of professional service for children is dependent upon cultural group membership. That can be demonstrated empirically. The failure to handle such matters as these, as psychological evaluation has not, in a sophisticated way, in evaluation studies, has led almost exclusively to a ghoulish psychological dissection of learners to locate the source of poor performance. The treatment, differential quality in the treatment that children receive, in general has escaped our attention. We have the technical capacity to remedy this injustice.

Finally, I will discuss cultural retardation among evaluators. Many phenomena in urban schools have been invisible to evaluators because of the general cultural retardation among them. As my brother, Thomas Hilliard, has shown, professional programs of preparation seldom, if ever, include the development of expertise in cultural skills. Therefore, evaluators arrive on the urban scene with its rich cultural mix and actually operate as if the cultures which they see there did not exist. They doggedly force the use of inappropriate tools and concepts on the phenomena and overlook the devastation that their own actions have created.

Elsewhere, I have tried to illustrate this matter with data on black language, one example of which is this. The ignorance of African language antecedents to the language that African-American populations speak will render certain standardized test data unintelligible or uninterpretable. Phonetic retention, for example, from Africa in the speech of African-American children is not treated on tests as retentions, but as pathological deviations from a EuropeanAmerican norm. Quite clearly, to a culturally skilled observer, such a norm is inappropriate. Yet, no normally trained educational evaluator would be aware of this since nothing in his or her training is likely to have provided the background for an awareness. Yet no amount of statistical evaluation design sophistication can correct for cultural retardation among evaluators.

The evaluation of urban programs in education must not proceed in the absence of evaluators who are trained not only in the normal methodology, but in the cultural background of those populations with whom they intend to work. Such cultural training must include history, language, and, in general, cultural information as well.

There is one last thing I would like to say. It concerns evaluating what. We have spent many years looking at and dissecting educational failure. In that same period of time, we have had outstanding educational successes, many of which are not funded programs. Where were we as educational evaluators when September Clark reduced the illiteracy rate in eleven southern states from 12 million people to 12,000, teaching reading? We were busy looking at IQ scores rather than at successful educational practices.

In conclusion, the specific recommendations that I have made are as follows: cultural preparation, specifically for evaluation professionals; better descriptions of the actual treatment that children receive in empirical terms; empirical definitions of content validity for criterion measures; empirical support for construct validation for criterion measures; more appropriate aggregation, particularly aggregation at least at the treatment level, for evaluation data; multidisciplinary methodologies rather than the exclusively psychological; the elimination of street language in professional terminology; and, finally, a focus in our evaluation on the successes in education.

PAT ALLEN, Teacher Corps: Dr. Hilliard, I am wondering if you would expand a bit on your comment about Piagetian observation related to empirical data. I would like to make a quantum leap then to Dr. Farrar's statement about process evaluation, whether that is not truly mainly a process observation, and maybe there is another step there. I don't know if there is some way for you to tie those in together or bring that out a little bit.

DR. HILLIARD: I guess I was appealing for empiricism for across the board, and empiricism and observation, not in the narrow sense that some people talk about, quantitative versus qualitative empiricism. You can have empiricism in both. The process observation could have and should have empirical backup. I was simply trying to cite, by using Piaget, in contrast to some of the other constructs that we work with, that his constructs derive from the observations with children, and most of these other constructs do not derive from that.

The same principle would be extended to a look at process. In other words, as I tried to say, I strongly support both the process and the outcomes, and they ought to be done simultaneously or else you cannot explain the outcome. But where we have been short -- I think that is why we invented the notion of formative evaluation and summative evalaution, to have a legitimate way of spending time only in one camp or the other, where we need to be in both simultaneously.

So I would go for the observation of the process. That is why I was talking about other disciplines necessary to do that. Basically, the kind of data that we normally collect from psychological disciplines are not process-focused at all. That is why you need participant observation, ethnomethodologies, and other kinds of things, to illuminate those processes. I really agree that is important.

GEORGE LOWE, Office of School Improvement: Dr. Allison in Chicago was discussing the cultural bias of tests 30 years ago. I was wondering what you think about why we have delayed so long when this thing was so obvious to us 30 years ago. Does it relate to the nature of class in society? What got in the way of it?

DR. HILLIARD: Both, the nature of class and society.

I was on a panel two nights ago at Georgia State and a local ETS person was there who just endorsed Jensen's book as the tombstone for discussions on test biases. He said Jensen's new book on test biases just answered all those questions. Jensen asserts that there is no such thing basically as test bias for most populations, that we all basically have the same environment, so there is no need to look at those kinds of things.

But, again, the problem hinges in part on a limited definition of test bias and a lack of empirical evidence on the part that they leave out. In other words, the limited definition is that they restrict discussions of test bias to information about item analyses, and if the item pattern response is the same for two different cultural groups, and one group is only lower than the other but the patterns are the same, the assumption is that there is no bias because of the statistical data.

But, for example, there are no appeals in those kinds of discussions to the linguistic data. So, for example, if you were asking a question what is vocabulary in the first place, even before you do your counting and statistical analysis, from a linguistic perspective it would be absurd what we have done for the last seventy years--measuring vocabulary. I think there are about 26 words on the WAIS that are vocabulary words. You are going to sample vocabulary with 26 words, and that does not even make sense in terms of what vocabulary is empirically form a linguist's perspective. Those are the kinds of things that never enter the discussion. So we still are stuck with it because of ignorance on the one hand and because, secondly, some people want it that way socially.

MARIE BARRY, Teacher Corps: Would you agree that in the field of psychology we are beginning to distinguish between the psychometrician and what I call the gestalt

psychologist which, in my terms, again, refer basically to the kinds of things that Piaget has made clear to us, and once you bridge the gap between the psychometrician and the gestalt psychologist, you do get into process and you can begin to build criteria such as you suggested in the linguistics and the logical fields that will give us another base for research and evaluation, including arithmetic once in a while.

DR. HILLIARD: I'm glad you brought that up. I agree fully. The only reason that I came down so hard on the psychological part is because those psychologists who are most closely associated with program evaluation are not the gestalt psychologists. They are from another crowd. Again, I am not even putting down that part of psychology because I think, quite clearly, that each part has some part to play. But the domination and exclusive participation and control by that crowd over evaluation I think is killing us.

MS. McKENZIE: Dr. Gold is Senior Research Associate at NIE and has conducted a number of evaluations of urban intervention programs.

EVALUATION METHODS APPROPRIATE FOR INTERVENTION PROGRAMS
BY NORMAN GOLD,
SENIOR RESEARCH ASSOCIATE
TEACHING AND LEARNING, NATIONAL INSTITUTE OF EDUCATION.

I want to spend just a moment changing the title of my presentation to "Evaluation Methods Appropriate for Intervention Programs," rather than "effective methods." The reason I want to do this is to make sure that there were not a set of expectations for what I have to say that I cannot meet.

By way of introduction and orientation, I was looking at a picture in "The Washington Post" a few years ago that really struck me and stuck in my mind for a long period of time, until this moment. It was a picture of a man on a sailboat, heading south, down the Potomac, east, then south, to the Caribbean. He had his family, dog, just about everybody on the boat with him and he was leaving. A reporter asked him, "Where are you going, and why?" He said, "I'm going to the Caribbean and am taking my family with me. The reason is because I either had to accept myself as a failure in terms of my set of experiences, or I had to redefine success."

It seems to me that with the literally thousands of urban intervention programs on the landscape, all of which have been determined as failures, there are possibly alternative explanations. We either have to accept the fact that all of the effort and all of the energy over the last 15 years is, in fact, a failure, or we have to redefine the measures that we used in terms of what we mean by failure

or success. It occurs to me that a possible explanation for this dismal record of intervention programs could be attributed to unrealistic expections of what can be achieved over a given period of time.

Each program that is formulated and supported for development promises a direct and impressive impact upon the set of problems for which the intervention strategy was defined. Students will perform better, will feel better and will act better as a direct result of the support experienced as participants in the program. These effects also will be sustained over time. Teachers and administrators, community participants will embrace the effort and continue implementing the program on their own. This effect will occur in some impressive degree within a specified period of time, and the program will be implemented, students will be processed through the program and will come out on the other side better.

Described in this somewhat irreverent way, it also occurs to me that one of the characteristics of these expectations for intervention programs is that they are somewhat similar to wishes. The problem as with most wishes, is that they often have limited contact with reality. This gets rather serious when we incorporate these wishes into a formal set of expectations, for example, in the form of government contracts and grants. If we have learned anything over the last 15 years, it is that the wishes and promises that initially gave us hope can come back to haunt us. We raise expectations of a wide range of people, including those we serve. We place excessive pressures on service providers and recipients. Probably one of the more tragic results is that we really have little opportunity for putting together all of the ingredients to mount significant locally-based intervention programs.

As far as I can see now, the opportunities will decrease for these kinds of programs over time. When expectations are not met, the individuals associated with the efforts lose credibility. The amount of investment individuals are then willing to make in the future diminish. It is my opinion that the disillusion and skepticism regarding social action intervention can be attributed partially to unmet and often unrealistic expectations. It has been my experience as well that the multitude of negative evaluations of urban intervention programs often reflect this discrepancy, between the wish for program effects and the reality of the complexity of the problems addressed.

I think most people would agree with the dismal record that I have mentioned. We either have to accept the fact that we have failed and these failures are associated with a continued need to create realistic expectations for intervention programs, or we must redefine success by accepting

more modest gains achieved over longer periods of investment. At this point in the development of evaluation, I think perhaps the most appropriate role evaluation can play in social intervention programs is to contribute its expertise in efforts to forumulate more realistic program expectations and implementation strategies. About five years ago, I started to take this issue seriously at the outset of evaluations, rather than as customarily done when the evaluation is completed. To assert that expectations for a program were unrealistic when it was over was a benediction that most of us were not very comfortable with.

I will now describe a set of operations and characteristics of evaluation to support the program development process. These are: modifying stakeholder expectations; developing inductive and dynamic evaluation strategies that are compatible with the development process; moderating expectations for evaluation; and measurement strategies for developmental programs.

Given the need of program developers and program funders to over-promise, it is likely that initial expectations will remain considerably optimistic. It is also true that our ability to predict at the outset either the course of program implementation or effect is still severely limited. Appropriate evaluation should assume a large share of the responsibility of providing information useful in grounding expectations. As with our friend the sailor, the adaptation process is to either change what you were doing to meet a set of expectations or to change your expectations. This process is necessitated by what is often interpreted as failure; but, in fact, it is gaining insight and experience. The evaluation process employed to aid in developing more realistic expectations is through successively approximating what is possible, given a set of resources, a period of time, and an increasingly elaborated conceptualization of the problem being faced.

A first task of evaluation is to negotiate an initial set of expectations that are fairly well rounded. This is done by piecing together representative expectancies from a variety of sources. Utilizing a variety of sources is important to avoid representing a single set of interests. The stakeholder group is designed to obtain expectations from all parties having a direct interest in the outcome of the evaluation. These groups represent federal, local, programmatic and participant interests. They are asked what they expect the program to produce and in what time frame. These sets of expectations are characterized by type and interest group.

The evaluation process then matches these expectations with both the proposed program strategy, the program development history, if any, and the history of efforts of this type. This provides some check on the reality of

expectations. The next is to determine what realistic form of evaluation perspective the existing programs can sustain: the capacity of the program, the time for the evaluation, the sources for evaluation, and the capacity of evaluation technology itself.

The initial evaluation approach emerging from this process as recommended to the stakeholders by the evaluation then consists of a set of expectations attached to a proposed or actual set of strategies that can be systematically observed during the life of the study. This proposal then is negotiated with all stakeholders and the evaluation proceeds only upon significant agreement on this or some acceptable alternative approach. This process can take from six to ten months to accomplish.

What is achieved is an evaluation based on a set of initial expectations that are somewhat rounded. More importantly, these expectations are shared in common over a range of vested interests and are viewed as preliminary and developmental.

There are many factors that will determine whether this initial expectation requires modification. Among the most important is the rate and extent of program implementation. The initial preoccupation of appropriate intervention program evaluation should be focused upon program implementation. Though some effort at obtaining base-line data for later impact assessment may be started, the major focus upon which all else will depend is the program implementation. It is often the case that program implementation will be more complex and take considerably longer than expected. It is also the case that to promote local involvement, considerable variation in program operation will occur. Finally, as knowledge of the complexity of the problem increases, it may be determined that expectations for behavioral change were too ambitious and more modest changes that proceed or function as proxies for more elaborate outcome measures may be more appropriate.

This type of evaluation ought to be a part of an annual process of reappraisal to determine if the rate and direction of the program being developed is as expected. If significant discrepancies exist between expectations and what is observed, then modifications in program strategies or goals and objectives ought to take place. This process of program modification should continue until the program stabilizes to a point where those concerned feel comfortable, confortable with the way the program is operating and with what is expected by the vested interests.

To enable this to occur requires explicit acknowledgement and support of an iterative process of program development from all parties directly involved. Though

this process in fact has operated in some programs, for example, in Headstart, it is enabled by political pressure and manipulation by program advocates. If more appropriate program development is to take place, serious rethinking of program development process needs to occur.

If we accept as at least a partial explanation of the observed problems in intervention program development the notion of reality testing and iteration, then our notion of the evaluative process must change dramatically when applied to this context. It is my view that the evaluative process in this kind of setting needs to be inductive and dynamic.

It is inductive in the sense that it gains knowledge through observing actual events. The event is the teacher. You learn what it is over time rather than matching the set of events against some predetermined set. The goal is to know the program as it actually occurs and to understand and describe as best you can why it has evolved in the way that it has.

It is the actual comparison between this process with expectations that are held by the population of stakeholders that allows the discrepancy notion to exist. It is not discrepancy as normally applied in evaluation. It is not discrepancy in terms of a pre-set notion of the program. It is discrepancy between expectations and what the program actually is, what has been observed.

The evaluation under these kinds of conditions needs, as well, to be dynamic, because you have to allow for the fact that programs can change and adapt rapidly and radically. If you set up a whole set of data collection structure, either it has to be so wide and expansive that it is almost too weighty to continue, or you have to be so specific that you miss a great deal of what the program actually is doing. Descriptive techniques, rather than large precept data bases, are much more powerful in this kind of environment.

The goals for evaluation need to be moderated. Unrealistic goals and expectations are not just the domain of program developers and program supporcers. We expect and have been told that evaluation will provide the definitive answers as to whether programs have been effective or not. Those answers also will be delivered within the time of the evaluation contract. The complaints about evaluations are that they are often equivocal, the findings are obscure, they are inappropriate for decision-making, they are poorly timed and they are not readily obtainable. This is the data which sort of gives us the feeling that there is a gap between expectations and reality, as far as evaluation is concerned.

So, what is reasonable to expect given the state of the art at present? What we can gain from evaluation is a greater understanding of the magnitude and nature of the problem. I think that has been very definitely and well described by Asa Hilliard. By understanding the problem better, by having better explanations of the problem, this can be important information to program developers and innovators, because they can begin to understand and help explain why they are having problems in certain kinds of areas, what they are missing, what they are not seeing, what kinds of adaptations or alternatives they might try.

We can use evaluation to aid in the program implementation process. Much of the time and effort of the evaluation people who work on projects that I fund are spent in understanding the implementation proces and providing information that is very useful in the iterative development of programmatic strategies. We can conduct analysis that determines under what conditions the program facilitates participants and in what way. This is very much in line with what Dr. Hilliard said in terms of sub-group analysis.

What we are not particularly interested in is great generalizations. All of our strategies have been designed in the past so that we can generalize what we are finding out. Then we find out we don't know anything and so we have nothing to generalize. The statistical techniques that we used, the particular designs are oriented toward generalization. In the program development context, this probably is not very useful. What we can do is to look and discover for the particular program at which we are looking where is it facilitating behavior and under what conditions.

Another problem is that we have to finally rid ourselves, so far as developmental programs are concerned, with the cost/benefit mentality. It is the cost/benefit mentality that keeps us having to find major effects.

If we can accept partial answers from several related, but independent, efforts, we can provide the foundation for more effective efforts in the future. Beyond the best case analysis procedures to which I have alluded are two ways cross study learning can be facilitated. The first is to break down parochial program interests, the notion that each program has to solve the major set of problems to get supported. A program should not have to promise total answers to complex problems, but is valuable if it can make a contribution to the knowledge base that we already have. Secondly, we have to continue to encourage cross program synthesis of programs that deal with the same population working with the same kind of problem.

Let me address some comments on appropriate measurement for intervention programs. First of all, it seems to me that we just ought to get rid of the set of things that

drives us in terms of generalizations. What we don't want to do is necessarily generalize, but we want to discover what works under what conditions. Programs have to be willing to be held responsible or accountable, that is true, but only for what they have some ability to manipulate directly. It seems to me that mostly they can be held accountable, for government purposes, to improve the provision of some service that they are providing. The measurement of this service is that it is useful to service providers and facilitators and it is attractive and useful to clients.

As far as finding out what happens to people as a result of the program, I think what we need to do is construct broad categories of behavioral changes that might be observed. Those broad categories ought to have in them a wide variety of possible behavioral indicators. I will give you an example. One set is a notion of investment behavior. Investment behavior is that set of behaviors which a program may affect which causes a student to be able to invest more in the educational experience. The sorts of programs with which I deal are very much oriented toward increasing this investment.

There are a whole set of indicators to show that, in fact, this investment is occurring. Attendance is investment. Doing work is an investment. Getting along in the particular social setting is an investment. Being less isolated is an investment. The point is that we don't know nor can we be responsible for the specific behaviors that people will adopt or modify.

I think one mistake that we have made is that we have had to promise that if we do this, these are the behaviors that will take place. I am concerned about that as a general issue. I am concerned that what we do is to place a conceptual and measurement straitjacket on service providers and recipients. If you are operating appropriately, then you will operate by doing these things, and if you are reacting appropriately, you will start doing these things. That's why I think what we really need to do is to have much broader categories with a range of possible kinds of effects that could occur and to take this measurement straitjacket off.

Wherever possible, use observation of behaviors rather than standardized tests. It seems to me that there are tremendous amounts of evidence that are building up all over the place and that have been building up for long periods of time showing that our penchant for scale measures just will not teach us very much. Given the kind of variability that I have described, observation seems to be much more powerful.

The type of evaluation procedures described in this presentation places evaluation squarely in the program development process. The notion of formative versus summative evaluation is obscured. Evaluation contributes in ways it knows how by testing discrepancies between expectations and observations. It also documents the history of the development of the programs. Both of these qualities are extremely useful for learning and contributing to the adjustment process. The strategy is to reduce the adversarial notion of both program intervention and its counterpart, program evaluation. The intervention is designed to facilitate local efforts at improving practice, not to supplant those efforts.

The goal of evaluation is to help in the program development process rather than judge. Judgement is held to a minimum. This is appropriate since evaluation, as we know it, at least at this point, never does very well at this job anyway. Final judgments usually are made on a whole array of factors in any event. Evaluation thus is cast more in a positivistic framework in an effort to contribute vigorously to the opportunities for more successful program development in the future.

SHIRLEY JACKSON, Director, Basic Skills Program: My question is for Asa Hilliard. It has to do with items discrimination as a vehicle for test bias. I want you to comment on this in the development of national standardized tests, the potential that is inherent in the whole process of items discrimination as a vehicle for test bias. I don't think that has been looked at too carefully.

DR. HILLIARD: Very briefly, item discrimination as a vehicle for test bias is usually duscussed most in the area of IQ testing as opposed to achievement testing. It is most appropriate, I think, to look more closely at it with IQ testing because it helps to determine whether or not the thinkings that are supposed to be measured, the behaviors that supposedly are being tapped, are really being tapped. But to limit the focus in that instance of the discussion on item discrimination and eliminate the rest of the information that we need to me is unscientific. It is unscientific because empirical data in other areas relevant to that question are completely left out.

Now, people also look at the patterns of groups on achievement tests, and there are some instances in which there are cultural variations. That is what I was trying to illustrate in the area of reading tests, in particular. Here the pattern analysis would not be sufficient to determine whether an item was biased. But what we finally determine we are going to do about it depends on the use of the achievement test. If the use of the achievement test is to measure whether people have come up to a norm upon which we have agreed, then I am less concerned about the issue of bias

in the achievement test. But if the achievement test is used in a diagnostic fashion, then the whole issue of bias becomes much more prevalent in my mind. I could say a lot about it. As a matter of fact, I have a whole chapter on it coming up in about two week.

GUY SHEFFLER, Teacher Corps: Dr. Hilliard, in your speech you made one remark which I would like you to clarify. You said that there were some teachers who got along with the troublemakers. My question is how do you clarify troublemakers in that fashion? Do you clarify it as being learning disabled, or how?

DR. HILLIARD: Well, I don't think I used the word "troublemaker." What I was referring to is that there are some children who are thought of as hard to teach. Usually we are talking about poor brown kids, poor white kids, and people who go into those communities frequently believe, and I think we see abundant evidence, that nothing much can be done for these children.

One of the reasons we believe that is because we do not look at what is being done under certain circumstances. In other words, I was saying that not all teachers fail with the children that we have identified as unteachable. When I say "not all," I mean that there are thousands, literally thousands of teachers whose performance is obscured by the way we collect information. This leaves us with the feeling that no one is doing anything because, on the average, not much seems to be done.

One of the ways I have tried to deal with this is to collect both anecdotal and program information on successful educational innovation. It is similar to some of the things that Ron Edmonds has done. I don't see how anyone can look at Edmonds' work or the work of some other people, such as James Comer who have seen things turned around with the unteachable, and can continue to have low expectations for the children rather than to focus their attention on the intervention that they are not receiving.

I don't even call the children troublemakers. I think troublemaking is a sympton of the fact that they are not getting attention. Those same children have long attention spans and are completely involved and need no external discipline when they watch TV. But there is a good show there. They don't have good shows where they are.

MS. McKENZIE: Thank you very much. You both gave us a lot of insight into appropriate methodologies for evaluating urban intervention programs. My point of view is that it is surprising that Norman and Asa were not not so very far apart in how they approached this topic.

DR. HILLIARD: I have to make one observation. If I understand the basic part of what Norman was saying, if it was focused on unrealistic expectations for some programs, in other words, if there are programs that really were set up to do limited things and we evaluate those programs as if they were supposed to be taking care of other things, then I don't have much problem with that. But if the unrealistic expectations are that programs should not be expected to change the outcomes for children, then I would have a lot of problems with that because if anything, I want evaluation to raise the expectation. This is why I gave the example I did in the beginning.

I don't see how we can sit here in America and look at the progress in the Soviet Union, look at the progress in Germany, look at the progress in Japan, look at what they do with urban educaton, and not realize questions about why we are so far behind. I think it is time now to jump anything that does not tell us how we can change that gap. I am never going to be satisfied with any explanation that suggests that it is in the genes of the kids. I have not been able to figure out how they could determine the genetic difference between Russians and Americans.

MARSHALL SCHMITT, National Diffusion Network: I also think we have to be careful about taking statistics from other countries and making a comparison with kids in this country in terms of proposed numbers of kids and the types of kids in the Soviet Union and elsewhere, comparing them to the general kids that we have.

DR. HILLIARD: Well, I have to say something on that. I know that this is something people sometime say. The way we ought to determine whether or not we can say that is by empirical observation. We don't have empirical observation of what happens when adequate treatment is provided to children. That's why I say we never have described adequate treatment.

We do have empirical information. I did not take the time to list all of these, but if I had to, I could. For instance, there is the Oakland Community School in Oakland, California; there is the Marcus Garvey School in Los Angeles; there is Marva Collins' school in Chicago. There are just any number of times that people have gotten fed up and have decided to go and do something different with these kids. In other words, there is nothing wrong with these kids. When we say that we suspect that there is, we are not dealing with empirical data. We are dealing with what happens to these kids when they don't get a good program. We are looking at a program which has been terrible for them.

MR. GOLD: Let me address this for a moment. I said nothing about expectations for individuals in my remarks and they should not be interpreted in that way. I really

said nothing about expectations for programs. What I said was that expectations had to be somehow related to what you are doing, the actual events, the actual set of operations that you undertake to make those changes. What my experience has been is that there is tremendous discrepancy between rhetoric and reality. My feeling is that it is not a very constructive role for evaluation to simply document that discrepancy. My studies are continually interpreted by the press as showing that things do not work, and tney keep reflecting this discrepancy.

What you then have the option of doing and what I think is an important contribution that evaluation can make to program development is you have the option either of changing what people expect or changing what you are doing so that you have a chance of meeting the expectations.

Is that clear? It might not have been clear a moment ago.

MS. McKENZIE: I think we will have to ponder it for a while. Let me make one footnote with respect to the comparison of our youngsters with the Soviets, West Germans, and Japanese in science and mathematics. There will be a report, I think issued from the White House, very soon on a study of that situation. I think we are going to have to look at it very carefully because basically we do have low expectations for our youngsters in those two areas of study. It will serve to our detriment and to their detriment because we are not keeping up with technology in our way of education and providing young people with a notion that learners should be dependent on educators rather than moving to push learners to be independent, using all the strategies available to them. So, with respect to poor youngsters and minority youngsters, and the youngsters that do not get adequate support from families or communities, we do need to push and to examine what we are doing with them, particularlly in the areas of science and mathematics.

PANEL DISCUSSION:
ESSENTIAL ELEMENTS FOR URBAN EDUCATION INTERVENTION PROGRAMS

INTRODUCTORY STATEMENT BY KATHLYN MOSES, DIRECTOR, URBAN INITIATIVES, OFFICE OF SCHOOL IMPROVEMENT

Each of the panelists is himself or herself a principal speaker, a keynote speaker. So to ask them to serve on a panel took a lot of "hutzpah." The next point is that you may look at this panel and think that you have here a Mulligan stew. You are absolutely correct. We have the press represented; we have research represented; we have a large public school system represented; we have a suburban school system represented; we have a classroom teacher represented; then, to really put the icing on the

cake, we have a recipient of all of our responsibilities, and that is a student, who will evaluate the program, not in a scientific way but from the point of view of what an urban intervention program has done for her. It may be different from any research you will read in papers, but this is the bottom line.

JAMES P. COMER
MAURICE FALK PROFESSOR OF CHILD PSYCHIATRY,
YALE CHILD STUDY CENTER,
AND ASSOCIATE DEAN, YALE UNIVERSITY MEDICAL SCHOOL

Ms. McKenzie had trouble defining what I do. I have trouble defining what I do as well. I do not consider myself a researcher, unless it is an intervention researcher, or an evaluator, and I kind of cringe when I am called such.

I think I got started in this business because of my personal experience in looking at the discrepancy between the performance of my low income classmate on the playground, in church, and elsewhere, seeing that they were obviously as bright and able as I was and others were, and yet that they did not perform well in school. Clearly something was wrong. That, in the long run, led to my concern about this issue. Also, becoming a clinician eventually and with a public health background, it seemed reasonable to be involved as an active participant-observer in the school program, because it appeared to me that there was an opportunity to apply the principles of the social and behavioral sciences to the problems in school.

We did exactly what Dr. Farrar said in our school program in New Haven, where there was a dramatic improvement in achievement and behavior among students in one of the lowest income schools in the city. They are now essentially at grade level in performance. We stepped in and messed around. We felt from the very beginning that the problem could not be with the students, because, in my heart, I knew from my own experience that they had the ability.

Clearly, my concern about formative evaluations being imposed and rigorous research projects being imposed on systems shows that there is so much that must be done before you can do any kind of evaluation in systems. We simply ignore it as researchers, generally. The intervenor must establish trust among the people involved in the system, and that can take anywhere from two to five years, depending on where you are, right off the bat. You have to deal with the dead wood in a system. A system that is down over time has a lot of dead wood, and you know it. To pretend that it is not there is dishonest. You have to go in and you have to deal with that dead wood and at least get it out of the way or get it to a point where it cannot interfere with the program. Your have to deal with opportunistic

people and manipulators. You have to deal with the press, which wants to destroy you. You have to deal with all of the doubters, all of the people who are there after you.

Now we know this. We cannot go into systems and pretend that we do not have to do all of those things. You have to do all of those things before you can permit a staff to systematically address the educational issues the treatment programs that you are going to provide for children in a program. That is part of the process of intervening in schools.

Also, the implementation of the program and looking at the implementation is important. It seems to me that in order to intervene in a school system and implement a program, there are several things you must do. You have to have reasonable intelligent intervenors in the first place. They don't have to be brilliant, just reasonably intelligent. You have to have people who have reasonably good interpersonal skills and relationship skills. I see people trying to intervene in schools who cannot get to first base with the people with whom they are dealing because they themselves do not have good interpersonal and good relationship skills. You have to have people who have some knowledge of human behavior, of system behavior, and of what Dr. Hilliard talked about, the cultural style and ways of the community.

But you also have to understand what is a by-product of trying to function and operate in a system that is essentially hostile to your goal and what kind of behavior that promotes. That is a large part of the problem. When I listen to people talk about parents who do not want to participate, a large part of that problem is related to the interactional problem between the groups who have been denied the system and closed out of it and the people who are in control of the system. You have to have that knowledge of the history and the relationships that Dr. Hilliard talked about.

Finally, you have to have people with good problem-solving and/or management skills, people who are capable of establishing goals, of establishing strategies and of developing mechanisms in which you can develop specific programs or treatment relative to the needs in that particular school. You need people who are able to measure the behavior changes and outcome changes within a school, who are able to identify the problems that they see as a result of the data they collect.

Let me give you one specific example to try to concretize what I am talking about. We have a learning center in our school program. The learning center people went in and worked with children who were supposed to be

retarded. They discovered after their program that the children who were supposed to be retarded were coming out scoring higher than the so-called "normal" children in the first and second grades of the school. Well, something obviously was wrong with the program being given to the so-called "normal" children if the others were out-performing them. If we had had rigid, separated, rigorous programs that did not look at the total school, we would not have found that outcome and we could not have pinpointed the problem and helped the staff to see that the problem was theirs and not the problem of the children.

In order to work in that way, it seems to me that evaluation has to take all of the issues that I have just mentioned into account -- the style of the intervenor, the staff's ability, the knowledge of human relations, system behavior, cultural and interactional problems. But, in addition to this, we have to consider as we evaluate programs the stability of the staff in terms of staying power. When you are always turning over, always starting over, then there is a very serious problem because you will never move forward.

The rate and direction of change, as mentioned by Dr. Gold, is tremendously important to look at as we do evaluations. There also are some other things about which we never talk; and yet, as policy-makers and researchers we know that they are tremendously important. These form the power issue, the control issue that go on at every level within a community, within a school, among staff, between staff and principal, between staff and support staff, at the central office level and at the state and national level. All of those issues are at play and we have to pay attention to them.

I agree very much with Dr. Hilliard's comment that in order to pay attention to those and other kinds of issues that are very troublesome, we need different ways of doing research. The ethnographic studies, the participant-observer approaches, and the case study methods used by business I think can contribute much more to our knowledge of school intervention programs than some of the kinds of approaches that we use right now.

Piaget was mentioned. I want to point out that Piaget was a clinician. Piaget asked question as a clinician would ask questions. One of the problems with our studies and one of the problems with our use of questionnaires is that when we ask questions, we don't know whether we are asking the right questions. We don't know whether the respondent is answering the question we thought he thinks we are asking. So it is very difficult. With the approach that Piaget uses, you can move from one question insight to another question. This is much more helpful that the approach we use right now.

I think that in order to look at something as complex as a public school system and the social system within an individua. school we have to use an ecological approach or an ecological perspective. We have to examine each element within the system -- parents, teachers, administrators, the young people, their training, their development, their capacity to do the job they say they are going to do or want to do, the kind of support they are receiving to do that particular job. Each element of the system must be looked at and not just the job.

I think that a second element of an ecological perspective would be that we have to look at the process and content of the input or, as Dr. Hilliard said, the treatment that the group uses. Is the intervention conceptualized correctly, based on what we know about the children, the school, the power issues, and so on? Have we prioritized what is important?

Someone mentioned that we have to look at the services. I think in a system that is chaotic, you can do almost anything and have some kind of good outcome. But is that good outcome what you really want? We have to say what is it that we want and is the treatment designed to bring about what we want; and, even if it brings about what we want, is that relevant to the major issue, which is improving the academic and the social performance of the young people. We have to ask that question. I personally feel that too many programs are funded to achieve frivolous goals, and those goals are achieved, but that is not essenially what we are after in trying to improve urban education.

Finally, do the intervenors have good timing and are they aware of timing and are they able to respond to the timing issues? Are they opportunistic? Are they aware? Do they watch for the changes in administration downtown, the changes in rules, the people who move to different positions who can be helpful and useful? Are they political, in other words? Now those may be dirty words to researchers, but those of us who are muddling around in a messy area are concerned about these kinds of things. The kinds of things we try to do in the first year of our program we could not do if the person in charge resented our being there. But when that person moved off, and sometimes they were helped to move off, we were able to do some other kinds of things. Those are the issues. That is the way school systems work, and we have to be responsive to that as we try to evaluate school programs.

Finally, the interaction of the various elements within a school system, the people who are involved there, parents, teachers, administrators and children, and how they interact all have to be looked at if we want to determine the effectiveness of a program. Interaction is tremendously important, and yet, we pay very little attention to it as we try to evaluate programs.

In conclusion, I would just like to say that the way we do evaluation research I think interferes with achieving the kind of knowledge that can be useful. I think that we have to take an approach that examines every aspect of a program and the way the elements of a program interact with each other if we are really to determine what needs to be done within a particular school system and whether we are able to pass that information that school people can use to change systems. All of the knowledge and all of the training in the world will be useless, in my opinion, if the people at the local level do not have the skills to use that knowledge in response to the kinds of opportunities they see at the local level. My plea to those of you who are funding evaluation programs is to have people pay attention to the kinds of issues that I am talking about here.

RAMON C. CORTINES
SUPERINTENDENT OF SCHOOLS, PASADENA, CALIFORNIA.

I am here to tell you that we can educate kids, whatever their status or lack of it, whatever their color, language, or abilities. As an article of faith and fact, I believe this to be true. We can educate kids and we can do so, whatever their backgrounds, despite those circumstances which statistically may tend to accompany poor educational performances and which are all too easily and too often used as excuses for it.

Since 1972, with one brief interruption, I have been Superintendent of the Pasadena Unified School District in California, a 23,000 student microcosm of a modern urban school district, one large enough to provide educational alternatives and small enough to manage them successfully, and to do this despite decreasing enrollment, increasing poverty and a rapidly changing ethnicity. For example, as the community has changed as a function of demographics, our enrollment has declined 9,000 students since 1966, a 28 percent decrease. Last year, close to 40 percent of of our students came from families receiving AFDC, and there is evidence that many families, particularly Hispanic families, are fearful of applying for it, even though they qualify. Nearly half of our student live with only one parent or guardian, and we have gone from a district which was 66 percent white and 24 percent black and 6 percent Hispanic in 1966 to one which last year was 32 percent white, 44 percent black, and 20 percent Hispanic.

In short, the Pasadena you see behind the Rose Bowl on New Year's Day, while not a painted backdrop, bears little resemblance to the community served by the schools. Although by geography we often are considered a suburb of Los Angeles, Pasadena, by history and circumstances, is a quaint, essentially urban community in its own right, and beset by the problems of nearly every American city.

From court ordered desegregation to the budgetary problems of recent years following Proposition 13, and in the face of inflation we have had to adjust our efforts and our programs to accommodate it, and I believe successfully so.

Let me illustrate with respect to our district's overall performance with respect to the performance of our compensatory education program funded by ESEA, Title I and its California equivalent. Since 1974, the district's educational performance, as measured by the comprehensive test of basic skills administered to all students each spring during the week in which the test was norm has increased dramatically from the 45th percentile in 1974-1975 to the 53rd percentile in our spring, 1979 test battery. Over the same period, the English language test scores of our Hispanic students increased from the 35th percentile to the 45th percentile, while the scores of our black students increased from the 27th percentile to the 42nd percentile. All of these gains, gains in terms or national norms, occured at a time when the district's population was becoming increasingly dissimilar to the norming group -- poorer, blacker, browner and all of the other excuses. Meanwhile our white population maintained its performance consistently between the 65th and 72nd percentile. Preliminary results for the 1980s indicates that this performance across ethnic groups has been maintained and overall the district is now at the 54th percentile.

Gains in our comp ed program, a program in which I have been deeply involved and of which I am most proud, have been perhaps our most impressive. For example, in a longitudinal study conducted by our planning department of research and evaluation, the Title I program was assessed in terms of academic performance of a matched set of program participants on the CTBS as they passed through our primary K-3 schools between 1975 and 1978. The results were examined in reading, language, and mathematics in terms of median percentile scores to make relative comparisons within the district between schools and against the national norms. Only those students who were tested in the district on all four tests, for example, each spring from 1975 through 1978, in at least one subject area, was included in study.

If we are to consider the fact that the test publishers' norms are based upon the assumption that all things remaining equal, a child's or school's percentile score is expected to remain the same in successive years, then increases in percentile scores represent greater than expected growth. In the case of the Title I program, such growth can be seen to represent evidence that all things have not remained equal, that, in fact, compensatory services have been provided and that the funds have had the effect they were intended to have.

The longitudinal study of the Pasadena comp ed program was, I believe, done as correctly as normal constraints allow. No effort was made to otherwise limit or define the data, to make the program look better. The study included all students within the district who had entered the school as kindergardeners in 1975 and who were still in the district in 1978, at the end of grade 3. Of the 824 students, out of over 1,800 who began in 1975, 640 were served in our 11 comp ed schools, while 184 were served in our nonfunded regular or fundamental schools. Almost two-thirds of the comp ed participants at the 11 funded sites were in the district's three lowest socio-economic categories, while approximately one-third of the students who would have qualified at the non-funded sites were found in the same three categories. In short, actual comp ed participants were, depending upon your viewpoint, half as rich or or twice as poor as their non-participating peers.

Now for the results. In reading the average comp ed participant gained 12 percentile points over the three school years, while moving from the 28th percentile in the spring in 1975 to the 40th percentile in spring in 1978. If this sounds unimpressive, you should remember that the method of analysis used discounts the inflated gains shown by the district which insists upon measuring gains from fall to spring. This growth is real, and I believe it represents a considerable achievement in the view of the frequency which we hear and read about, the declines in student performance across the nation.

"Pasadena schools," to quote the study, "can be seen to be successfully narrowing the gap between underachievers and the national, but mythical, average student, and even more so in language and mathematics than in reading." For example, in language the average comp ed participant gained 21 percentile points while moving from the 24th to the 45th percentile. In mathematics the average participant gained 25 percentile points while moving from the 22nd to the 47th percentile. Given the nature of kids, I find these gains impressive.

Let me illustrate them in another fashion. As schools, including all students enrolled, whether or not served by comp ed, the eleven comp ed schools outgained the more affluent nonfunded schools in all three subject areas between 1975 and 1978. They did so by 5 percentile points in reading, while moving as a group from the 43rd to the 61st percentile, an overall gain of 18 percentile points, while the nonfunded schools gained 13 percentile points. In language they outgained the nonfunded schools 4 percentile points, while moving from the 50th to the 62nd percentile.

While the evidence of success is based upon standardized testing programs, the district also uses various criterion

reference instruments, particularly at the K-6 level, and has developed its own proficiency or competency test for use at the secondary level to determine graduation standards. We have used our criterion reference tests appropriately and extensively to insure that what is being taught is being learned, and to place students at their correct instructional level regardless of grade level.

We have used our proficiency exams to insure that graduation from the Pasadena schools means something. We did it before the law required us to, and we are doing it with higher standards than typically are being applied. For example, when first tested in 1977-1978, 50 percent of our sophomores failed our proficiency tests. But, instead of lowering our standards for graduation, which would have been easier politically, we instituted remediation programs at the secondary level. In 1979, 61 percent of the students taking the exam for the first time passed.

By the way, that is an interesting study which I would be willing to share with you. It points out what happened to the students the second time, the third time, and the fourth time around the percentage continues to decrease. We are confident that these results will continue to show growth, that Pasadena's diplomas will continue to mean something, and that eventually, before they leave high school all students can be proficient in basic skills.

The point of all this testing, however, is not found in the results per se. Rather, their use in instruction and in communicating with the public is even of much greater importance. Each year in our district and our federal-state evaluation and test reports, we fully dissect the data for the purposes of improving instruction, for the purposes of revising and strengthening the curriculum, for the purposes of allocating funds and other resources. We do so publicly, by school, by grade, by program, although we have been extremely careful never to use our results for personnel evaluation. The fact that our results have been made public has proven to be quite an incentive for school improvement and for improving the reputation of the schools, sometimes if only for honesty within the community.

Getting back to and on the urban intervention and evaluation, however, I am not sure what our experiences in Pasadena mean except that they are our experiences and, given the choice, if any intervention is to be done, I believe it ought to be done locally and evaluated locally rather than prescribed by the federal government or the state. While there are a wide variety of successful urban educational programs around the country, they are successful as a function of local involvement, development, implementation and evaluation. I do not believe there are common elements or attributes; if they can be defined or divined, they can be formulated

to provide more than a common sense understanding of what ought to be done locally to meet the needs of kids, of school, of their communities. I am not, however, pessimistic, except with regard to the ability of the state and the federal policy makers to impose yet another program upon all of us at the local level as a function of what you may discover about some of us during a forum such as this.

There are a number of lessons inherent in any successful program. The question is whether they can be mandated for application in other districts, in other circumstances. Before concluding my remarks, I would like to discuss briefly what I believe to be the lessons of Pasadena.

One, there is a wide variety of instructional approaches, of programmatic approches, which work and work well, depending upon management skills. Between the either/or situations presented to us by those who see the answer is ever more individualization and those who would have us see their answer in terms of a slogan, in terms of back to basics, there are an infinite variety of teaching methods and programs practices which work. Acceptance of this fact constitutes an acceptance of reality. Despite the crusades of the reformers and the mandates of reform programs, success remains a function of local application and not philosophy or legislation.

Two, urban kids, any kids, do not need special education, with apologies to the advocates of 94-142. They need a good education. The difference between the two is more than semantic. Special programs, whatever their content, are too easily seen as appendages to regular instruction rather than as an integral part of it. Special programs are too easily seen as someone else's responsibility. Whatever the nature of the program's content, a school administrator must be made to be and feel as responsible for it as he or she is for the so called regular programs or curriculum.

Three, everyone within a school, a district, a community, can or should, whenever possible, be seen in the broadest sense as an educator. This includes involvement of their parents, although too many educators confuse attendance at meetings at back to school nights, in resource rooms, with a kind of day to day involvement in their child's education, in their child's progress, that I feel is considerably more important. Further, when I say everyone in the system, I mean just that. I mean that those procedures which are developed to comply with the dictates of the law or the need of management should be developed to support those services and staff members which are more closely involved with kids. Too often, requirements at the education center, and I refer to my own office as "Puzzle Palace," at the district level are allowed to supercede those activities which are designed for kids.

Despite regulations and dictates of the administration, I have found that the administrators can learn to provide services, supportive services and products rather than merely adding to the burdens already placed upon the rest of the system. In Pasadena, such supportive services and products have led over the years to higher standards of program design and planning than typically are required, to on-site assistance for student identification, to useful testing and evaluation reports, to useful on-site review and assistance and much more in terms of curriculum planning, staff development, implementation and evaluation.

Four, none of these lessons applies unless they are accompanied by an honest and an open willingness to undertake the challenges of public education, of educating kids, a commitment to see the task through to its end each year with a new set of kids, with the same enthusiasm that kids as kids can bring to the process, and a belief that it can be done, that kids can be educated by us. There is neither secret nor panacea in which I believe to be Pasadena's success, and there are many problems that remain. But I do believe we can have an effect on kids, that we can educate kids for the better of us all.

We can educate kids in the years and the battles ahead. I would hope that the defense of public education can rest upon our successes, our products, rather than, as all too often has been the case, merely upon our hopes and our promises. There is much to be done and much to do within the system and within the budget, without succumbing to the illusions of reform and the enticement of new funding, or the temptation to so narrowly define proficiency that subsequent claims of successes belie our broader efforts, the expectations of the public, and the needs of kids.

NANCY HICKS,
PRESIDENT,
THE INSTITUTE FOR JOURNALISM EDUCATION.

I am a journalist, a sometimes educator. I have the feeling that with all the heat I had to take off my armor; it is armor that I put on when it was explained to me that my role would be here today was to explain how come the press kills everything before it gets started.

MS. HICKS: I will try to deal with that. I first wanted to explain to you a little bit about how I come to these issues because it will affect what I have to say about them.

I am a product of early intervention in urban education in this form. In the sixth grade, in 1956, in New York City, my class at P.S. 46 was intervisited with

a white class in our district. After the Brown Decision, there was some feeling that we needed to get to know the white kids. I don't know that what we were supposed to get to learn, but I think my most lasting memory of that is that we got served agricultural surplus in the lunchroom and they did not.

I have been a working journalist for 14 years, four of those years as an education reporter with the "New York Post" and the "New York Times," cutting my teeth on the Ocean Hill-Brownsville school decentralization issue, logging more hours than any journalist living or dead covering that story.

Right now, I am running a journalistic teaching institute called the Institute for Journalism Education. We have training and placement programs for minority journalists. I guess we have kind of a professional intervention program that has many of the imperatives and problems of yours in mass education. We are nine journalists who took a covenant. We looked around our profession which, after ten years of a lot of hard work, still has only 4.95 percent professionals of color in the ranks of daily newspapers.

We could have taken a structural approach or we could begin etching away at the problem, doing it a little bit at a time. What we did was to set up first a program to train entry-level reporters and to place them in newspaper jobs. The industry said there was no one qualified. So we said "We will give you someone," and we did that. Then it looked as if we needed more managers, so we set up a program to train experienced reporters to be managers and put them in newspaper jobs. Well, we didn't need to train everybody. Everyone does not need education. So we set up a job referral service that operates here in Washington so that people who needed no training could have jobs. The industry could not say that there was no help in finding qualified professionals.

What we did with the industry was we gave them the ball back. They said we are not qualified the parallel of which is kids cannot learn. We said here, you have not welcomed minority professionals into this business. That's the problem. It's your fault, it's not ours. What are you going to do about it?

The success rate has been quite something for the programs that we have. We have about a 75 percent retention rate in the business after three years, and this is a business that has a 40 percent attrition rate in two years. So we are doing twice as well as the average of black or white reporters in the business.

When you look at intervention programs, I think you don't see that assumption in the school system, that the job is theirs to do and not the kids'. I was glad to

look at the literature for this program and see that that is turning around somewhat. The accountability of teachers and educators is becoming more the norm than not.

But, for a large measure, a lot of the programs are oriented to behavior and not to learning. While that may make for orderly classrooms and there may be some better results on test scores, I can tell you what it means by the time students who have been in programs like this get to meet. What it means is that their drive and their spirit is taken out. We see more and more competent young folks who have no daring and no ambition and no ability to make the system work for them. I think as educators that we have to think as much about that as about the test scores. What are we doing in making nice, neat programs and test scores if children are not allowed to disagree or to come up with their own formulations of doing things?

I was very impressed in looking at the Larkin study out of Milwaukee for this reason. We have paid attention, right-thinking people on this issue, to the idea that every child is valuable. Well, every educator or every teacher who is in position and who can't be moved also is valuable. I you have someone in the classroom that the union says you cannot move, you'd better figure out how to make that person work for you. Hers was the first data I had seen on a program that went directly to trying to change the attitudes and performance of teachers who are in place and who you cannot move. You are not always going to be able to sidestep them.

In our programs we are very lucky as we can select broadly and get who we want. You understand that every once in a while you are going to place a person in a job, for example, with an editor who is not very friendly to him or her, and you have to deal with that.

All of this plays into the coverage a little bit. The reason we are in this business, what some people call social experimentation, is to change the portrayal of non-white people in the newspapers. This is a sort of tail chasing the dog around. Part of the problem of expectation of educators comes out of what they read about minority people in the newspapers or what they see on television. They seem to see that we are less intelligent, less hard-working, less able to concentrate. This is a function of having a cultrually biased press, which is predominantly white and middle class, in a country that is about 20 percent minority, as I think the latest census will show. They say that we don't know English, and this has kept us from getting jobs in the business and being able to portray a different life than that which you see every day.

I would like to take a little look now at what the press, biased and unbiased, meets when it gets into the educational coverage area, an admittedly flawed press.

I'm sorry to say that I have found in the range of coverage that I have done in my professional career that educators are among the thinnest skinned folks that I've run into. They are less straightforward, and I don't mean that they are not honest. But when someone asks you a question about how your kids are doing, the answer can't be, "Uhh." You've got to have an answer. If you want to know what that looks like to someone who is interviewing you, play it out on videotape once. Every pause looks like indecision and uncertainty and like you don't know what you are doing.

And, there is a naivete among the ranks of school administrators, given the political climate in which they are dealing. Politicians think of the press as parents. We give birth to them. Educators think of us as assassins. Yet the public educators are as political and they should recognize themselves as such. In order to obtain the financial and other support needed to have an intervention of any kind, you already had to sell someone, and you may or may not know is going to exist. But the expectation is there for success. Part of it is your own problem. This, again, is why the Larkin program looked very attractive. They did not sell anyone on anything in terms of more money. They tried to reshuffle the deck and come up with something with existing resources.

One suggestion I might make is to look at where you go for your experimentation money. Try to use your friends for that, your old friends. I know that we do that. When we want to do something that we are not sure is going to work, we go to people who know us well and who are willing to invest, to risk capital in our program development. We use new sources of support for things that are tried and true.

If you are going to try something that you don't know will work, if you have an alternative, don't go into the public-political market place for money. Try some other place. I think that might change the climate of expectation for your programs.

I know that research is careful and that it takes time, but you have to have a way to have a short-term assessment of what it is that you are doing. I've read all of the papers in the packet, and I've not seen one number. You need something concrete. The logicians tell us that things that are concrete are proper names, addresses, and numbers. If you don't have a way of at least putting markers on some of the kids and knowing what you want to show, you cannot put off the public for that long. There are the politicians who have the budgets to approve. There are also the parents who are offering their kids to your experiment. You have to be accountable in the short range to those particular constituencies. You cannot be euphemistic. I've never seen a newspaper lead include the words "interface," or

"empirical" or "mention" or "fell." There are hard verbs. You have to think in terms of what it is that the public looks at. The soft pause, the "Hmmm, let me think about it," does not go over well in the public arena. It may be fine in an arena of peers, but it is not fine anywhere else.

I don't think that we are off the hook in regard to coverage, but let me try to put together a short list of things that you can do for yourselves and then bring to us. Number one I would think is to know what your are trying to do and how to measure it. Ask yourself the question, at least in a way that sounds human and public. It may not be in the jargon in which you are used to working. But do think that there are other people who are going to have to support this.

Pick your principals carefully, and, if you can't pick them, train them. If you can, have some contractual relationship with them over the results of your experiment. Do so if you can. I understand that this is a sensitive issue with unions and the like. But there might be a way if you think about it, of putting people under a real or implied contract to the results that you want.

Think of the children, of empowering children, not of their parents, because they ultimately can do it, even when we can't. I am a parent. My poor kid shakes his head and suffers every day about what it is that we do to his life. But you have to focus on them because they are the actors. The parents are not the actors in all of this.

You have to figure out how to motivate your teachers or to get your teachers in a good economy. I think school systems will do very well now in finding talented people to teach school. But part of that reason is that the job market is so tight, there are people turning to teaching now who would not at other points. I think my observation is that more teachers leave because they do not have the support of the administration, rather than the difficulties that they have with any individual children.

So structurally, in terms of your own resources and the human capital with which you are dealing, see if you can work in that area. Then, what do you do with this press that won't do what you want it to do? I must admit that you can't take my advice because for all of the programs that I've run, I can't get any press coverage either. Even though my husband is a newspaper editor, that does not help.

Before the problem comes up, what you have to is find a source on your local newspaper who you know and to whom talk. Develop some sense of trust with that person.

I think that to feel that he or she knows that you are honest before the crunch comes so that there will be some expectation of how to deal when there is a real problem. Be very clear about what it is that you are saying and try to summarize the main point of what your are doing.

When a story appears and you don't like it, look at the why of it and try to be somewhat detached. If it is a little problem, tell the reporter not to make a big stink about it. "I didn't like this. I thought you were unfair. Do it this way next time." If it's serious, then then you have a whole range of options. You can write a letter to the editor and it will be published. You can You can possibly write an op-ed piece for your own newspaper. Many newspapers more and more are offering their opinion pages to local folks to tell their side of things. If the paper has an ombudsman, and many are doing that now, you can appeal your case there as well. If it is a fatal problem, one that you feel has disrupted your ability to raise money or to nelp your children, there is a court of last resort in the newspaper industry now called the National News Council. It's located at Columbia University and it does put particular newspapers and television stations on trial for misdeeds against the public.

KAREN LAMOUREAUX,
LANGUAGE ARTS TEACHER
PARK JUNIOR HIGH SCHOOL, CHATTANOOGA, TENNESSEE.

This morning, when we got here and I listened to the rhetoric of the first few speeches, I began to feel a little out of place and wonder whether I really belonged here after all. But the more people talked, the more things came out to show me that we have a great deal in common.

For instance, increased planning time was brought up by Dr. Farrar when she was speaking of the evolutionary planning approach. We need this so badly. When we are dealing with federal programs, we are required to do more work, but nobody ever thinks about giving us more time to do it in. So I really appreciated hearing that.

Technical assistance in introducing new programs locally is so important. Just because you are local does not make you an expert on organization.

Dr. Hilliard's remarks about cultural retardation in evaluators I appreciated greatly. I think that a teacher these days has to be a cultural expert. Anybody who deals with children in a public system needs to have a really honest and open approach to the cultural problems that we face. I've been thinking this for ten years and I'm really glad to hear it brought out.

Dr. Gold's approach to realistic evaluation also is appreciated. He brought up the fact that if you have more than one program operating in a school, these programs should be tailored to meet each other's needs and to fit each other. Two programs that are opposing in the same school are horrible. It's just terrible to try to live through it. They play against each other. Teachers suffer, students suffer, and we all lose. This is very important.

I was asked to sort of state some of our needs. Just briefly, we need program continuity. We don't need a program that comes in and gets the ax in two years. It takes five years, from what I've seen, for a program really to work. We need time to permit programs to evolve to meet the true needs of the community. When a program is interrupted to be appraised or evaluated continually, then the effectiveness of the program is destroyed.

We need program synchronization. If we have, as we do, Title I, ESEA and PUSH all in the same school, it can be a problem unless the programs are tailored to fit. I can't tell you how important this is. When your objectives clash, then you really are hurting all of your programs and you are demoralizing your teaching staff in a way that you would not believe. Also, the children suffer as well.

I have heard the terms "researcher" and "evaluator" repeatedly this morning, and a lot of you are professionals in that area. I am not, by title, but I think I could be called a pure researcher because I deal with the real thing, right down on ground level. My real thing is right here with me, and I would like to bring her up at this time so that she can impress you the way she impresses me.

YOLANDA BYARS,
NINTH GRADE HONOR STUDENT
PARK JUNIOR HIGH SCHOOL, CHATTANOOGA, TENNESSEE.

I would like to tell you about PUSH/EXCEL in action. PUSH/EXCEL at Alton Park gives students a reason to attend school regularly. What I mean is that it encourages students like me to try to better our education.

The program for pupils with perfect attendance increases the number of students in school. The Paper Action, the Coke Parties, Hall Monitoring, the Governance Committee, and the Media Committee and Radio Goes to School are all part of PUSH/EXCEL. It is making students push to excel. It was founded by the Reverend Jesse L. Jackson to promote excellence in education and was designed for all pupils regardless of race, religion and success in school.

1979-80 was the first year. The objectives were to improve self-discipline, study habits, and attendance. Attendance was improved by 1 percent, and the activities and awards were coke parties, trophies, certificates, movies, small dances, Speak Out for Excellence contest, the News Awareness contest, the Who's Who contest, and Radio Goes to School.

The objectives for 1980-81 are to improve self-discipline and parental involvement. Regarding parental involvement, parents are given certificates every time that they attend school activities. Awards will be made to the one with the most certificates at the end of the year. Parents volunteer to supervise lunch lines, games, hall monitors and so on. Parents are involved by Happy Grams, newsletters, monthly newspapers, the Bridge Newspaper and the Radio Goes to School.

I'd like to tell you about Radio Goes to School. I am a former member of Radio Goes to School. The program is designed for students who are interested in radio. Students are auditioned by reading scripts. Ms. Karen Lamoureaux is in charge of it and she is also the faculty advisor of the program. Radio Goes to School also motivates the students. It is another reason for the increasing of the self discipline program. There are ten bright students on the program, which takes place every six weeks because of the other schools that are involved. The program informs parents, gives students the experience of working and speaking on radio and tells or talks about activities in and around school.

During my three years at Alton Park, I have tried to live up to the standards that I have set for my life. My achievements include being on PUSH/EXCEL's Governance Committee. It is a responsibility of the Governance Committee to determine that our objectives are identifying impediments to education. The Committee also is responsible for outlining activities intended to eliminate or reduce these impediments to education that I identified. Being a star-roll student in science, I maintained an "A" in the first and second semesters of eighth grade. Being an honor roll student in the seventh grade, I made the honor roll two times. In eighth grade I made it three times, and in ninth grade, I hope to make it all four times. I won an essay contest, coming in first place, on the subject of "Why I Like Chattanooga Public Schools." I am also the head majorette and the co-head cheerleader at Alton Park. To fulfill my needs of writing, I am writing for the "Bridge" newspaper. Radio Goes to School is also my favorite activity. It gives me a great experience working with radio.

MS. LAMOUREAUX: I think you'll agree that she's quite a little girl, just from what you've been able to see so far.

What I have to say now is not so happy and not so nice. The real reason that I brought Yolanda here is as a graphic aid, a visual aid. I want you to see what an inner city student can be, can do. Then I want you to think about this. Any number of her peers, kids who operate on her level, who produce with the kind of flair she produces with, who have the kind of composure, creativity, and imagination that she has, are just not going to make it. This student, hopefully, I pray, will make it. My point is that things happen to these kids because of their environment which destroys their chances.

In the audience right now is an essay which I have passed around. It is written by a young lady who is not so fortunate as Yolanda. This young lady, you will see from her essay, has creativity, imagination, intelligence and she is Yolanda's age. She is very attractive. When she read that in front of her English class, a number of people actually cried, it was so good. She is pregnant. Her chances for success are practically nil. If it happened to her, it can happen to any child in the inner city community. This is the community that these programs are structured to serve.

So, when I asked for program continuity, I meant it in several ways. First of all, PUSH/EXCEL in our school and in our community is an effort to promote value. It is an effort to teach children the idea of excellent, but also to give them a sense of value about themselves and values, period. Now they can come from home, but there are so many things in the community that work to destroy them that they need extra help to keep them up. PUSH brings in the churches. Where is a better place to go for value. It brings parents into the school. It brings up the expectations of the children themselves. So, please, do not cut it off because their social advances and achievements are in some ways much more important than academic achievement.

The little girl who wrote the essay is gone. Her chances are over. You can whitewash it all you want, but really they are over. Her intellect has nothing to do with it. The community did it to her. Also, in this area we have to look at the other part of this combination, and that is the little boy who fathered the child. Generally, when this happens the child who fathers the baby is a non-achiever or a low achiever. He might be older, he might not. The fact is that he probably reads on a very low level. He either does not have a job or can't get a job. He might have been to jail. He has any number of problems. In other words, he has no sense of value except as a father of a child.

In this regard, we need to think seriously about continuance of Title I programs, programs that teach kids to read and which raise their reading level. This is most

important. I can't tell you how important it is for a young male. They need this very, very badly, because they cannot get jobs. They can't compete unless they have this background. If this is a letdown, I'm sorry. You've just seen what a beautiful job Yolanda did. But if you don't see for real with your eyes and understand that it is not just the kid who reads on 1.0, it is a child like this who also can be just chopped down.

We need academic and we need social suport. In the communities where these programs provide assistance like this, help is going to come from no other quarter. It is up to you. I will close with this. I do not see anything wrong with looking to the Federal Government to provide social conscience. That is what government is for -- it always has been and always will be. I make no apologies for that. Please, let's keep these programs going strong for these kids.

STUART RANKIN
ASSISTANT SUPERINTENDENT FOR RESEARCH, PLANNING, AND EVALUATION, DETROIT PUBLIC SCHOOLS, DETROIT, MICHIGAN

I want to talk about evaluation as intervention instead of evaluation of intervention and will do so in a minute. But first I had better tell you from whence I come.

My present biases, and those of many of the intervention strategies that are going on in Detroit public schools, are out of the research of Ron Edmonds, Larry Lezotte, Wilbur Brookover, Benjamin Bloom, Torsten Husen, Rutter, Popham, Tidwell and Courtney Casden and Lebov, a couple of psycholinguists, and Dave, who knows about parent behavior, and Feuerstein, a follower of Piaget, who has extended some of that work into some other areas.

If I were to talk about what intervention programs I think will work in urban areas and seem to be working, I would talk about the following characteristics. The first one would be that change probably happens school by school; that is, the single school is the largest unit of change and maybe we ought to act on that when we are working on intervention programs. This means looking at the school as a social system; and the works of Lezotte, Brookover, Rutter, and Edmonds have done that.

The second characteristic is that a good program is going to have ownership; this means by all the key people in and around the program, and you know who they are. You have to work awfully hard in your involvement strategies, your communication strategies, and your implementation strategies to be sure that there is ownership by the people who have to make the program work. It probably would be a better idea if they were designed in the first place by the user, and then you would not have to

worry so much about whether they can buy into the program. I can get awfully excited about stuff I designed. It is a little tougher for me to get high on somebody else's design.

The third one is the one you have been hearing about all morning, which is that teachers have to believe that kids can learn. I think along with this I would want to say that we can have an alternative to the normal curve. We don't have to continue that forever. It's possible that we could have equity in education and that we could have it in output terms, and the distribution would look the same for rich and poor, black and white, male and female. It seems to me that now this is possible. I think we know enough to do that. I don't think that you can convince teachers to believe that kids can learn by the Rosenthal-Jacobsoncon. I don't think you can do it by telling them that they ought to read the research that says that it's possible, Edmonds' or anybody else's. I think the only way you are going to do it is to have them conduct themselves in the classroom so that they get the feedback themselves that those kids can learn. I think it is possible to surround them with the support strategies that will produce that. If you are interested in attitude change, I think you had better go after behavior change and see if, through that, you can get the kind of return and feedback that maybe will have an impact on attitude, but that will at least reinforce the behavior change.

The fourth thing is that teachers have to believe they can teach. I think it is almost harder to convince them of that than it is that kids can learn, although these obviously are closely related. But it is a terribly important thing to believe in if you are going to teach. (By the way, these items are not in any particular order.)

The fifth thing is that none of the studies have found a school that is successful where the principal is not the instructional leader. Now you may be able to. I have not seen one yet. The principal sees his or her primary task as improving the performance of staff and leading the instructional program.

A sixth characteristic is one that someone said earlier which I would echo. It concerns the way the parents get involved and support education. This is in parent behavior in the home regarding their own child, rather than as massive participation in all kinds of city-wide committees and so on. Now I am not against the latter; I just don't see the evidence of that making a difference.

Then there are four that get more into classroom behavior. This next is the hardest one to sell, or so I found in Detroit. Maybe it would be easier in other places. I think we need less differentiated instruction. We worked so darn hard at individualizing instruction some 15 or 20 years ago

that we got fairly good at it. We told teachers that if they did a good job of teaching, they would spread the distribution. Sure enough, in all those individualized programs of the 1960s, the rich got richer, the poor got poorer, and the mean stayed where it was.

There is another thing that has come along more recently to reinforce that, engaged time on task, that you read in the California Beginning Teacher Study. It seems to show that you should choose to have the student actively thinking and engaged in the learning task and tied into it. In other words, if you can either have readiness or you can have engaged time on task, it looks to me as if the weight is on the engaged time on task. I don't think the story is all in on that. I would be interested in seeing some more research.

This means that some of our grouping practices, not only those of between class grouping but also the within class grouping, ought to be challenged. Another thing is that there is nothing so terribly wrong about having a youngster practice the task in the classroom that you really are expecting him to learn This means that you had better define what that looks like in a measurement way early-on. There has to be some feedback, too.

Those are the kinds of things, it seems to me, as I look at the research, that we now know. I think ten years ago, if we claimed we don't know how to teach urban kids, we might have had a better argument. But I think now we certainly have enough information so that the excuse ought not be there now.

Now, what about the Detroit program? Now I am working with Edmonds and Lezotte to try to put some of that research that they have been doing recently into operation in Detroit schools. We have a lot of activity going on there. That is school by school. What kind of support structure do you need?

The program that I want to mention is one that James Popham and I put together over the last three or four years. It is called the Detroit High School Proficiency Program. It is that terrible thing of hanging a rap on the kids for the ills of society, schools, and homes, by having them have to pass an examination in order to get a high school diploma. We are doing it. We don't really withhold the diploma. We just endorse the diploma of those who do pass the test.

We gained the power commitment from the superintendent and the board early on. We have taken the trouble to involve the community appropriately, and all staff and youngsters and employers to help determine what the objectives should be. I think we have done a good job of defining the test specifications carefully. An objective for us is not a one-liner.

Do you remember those behavioral objectives that we wrote in the 1960s and early 1970s? We finally broke down the curriculum so that it was so small that it was trivial, it was non-additive. It was multiplying a two place number by a three place number, and consonant blends, and things of that kind. So, having done that, we ended up with 347 objectives for an elementary reading program. I would defy you to find me a teacher who could teach 347 objectives in one term.

We should move from norm referenced tests to criterion referenced tests because norm referenced tests are standardized by picking items that discriminate and items that discriminate are likely to discriminate because there are things you are not teaching or because they are biased in some way. So, therefore, if you do a good job of standardizing the test you can come up with an instrument that measures what you are not teaching and which does so in a biased way. That is over-stating the case, of course. I think far more bias enters the testing situation by the person giving the test or before you get ready for the test. So we have to do a lot of work there, too.

In any event, it is clear to me that defining very carefully what it is that you are expected to learn is important. In the first place, you pick something that is big enough, like being able to write a paragraph or being able to get the main idea out of a reading passage, or being able to use formulas in mathematics, something like that. You pick a chunk of curriculum which is large enough so that it is meaningful and small enough so that it is not still a goal, so that it really can be objective. Then you define the devil out of it. This means test specifications. It means very specific statements about what the stimulus in those test questions will be, what criteria the responses have to meet, what the subskills are, what some teaching strategies are, and then you lay all of that out so that you have a full instructional program.

Now you have done the job of analyzing what it means to learn how to write a paragraph. In Detroit, our high school English teachers are now telling us that within the last year they finally have learned what it means to teach paragraph writing. They have learned it because we did a pretty good job of defining what that looks like, we are clearly measuring it, and the chips are down. Incidentally, it does take a little time to learn how to score 30,000 or 40,000 paragraphs and to do it in a fashion that is fairly reliable. We have learned a bit about that.

Good learning is happening because the attention to appropriate measurement was given, because the objectives are few in number and very well defined and very important, because there is involvement and commitment and ownership, because the stakes are high, because there is a full support program of

in-service education, materials, and all of that to go with it, because there are practice tests, because it has been fed into the curriculum K-12 not only in English and math, but in other areas, because there is a remedial program in secondary schools, because we are using what we know about peer grcup tutoring and cross-age tutoring, and because we are supporting the program.

I think it will work. I don't know yet. We gave the test last January for the first time to our tenth and eleventh graders. 82 percent of the youngsters passed the reading portion at that point; 55 percent passed the writing portion and 50 percent passed the math portion. As you know, most of us have been emmphasizing reading more than we have those other two areas in our schools in recent years.

If we are talking about evaluation of intervention programs, or even evaluation as an intervention program, then we need to look at the beginning of the evaluation process and the end of the evaluation process more carefully than we used to. The middle is the data collection, the data organization and the data analysis and so on. The beginning is the focusing, the designing, the who's going to make what decisions on the basis of this, when are they going to make them, what are the objectives, what does this really mean operationally and so on. The end is that the evaluator does a decent job of not just handing somebody the report, but seeing that the report is interpreted, understood, and used. I think we need to work at those two ends as evaluators if we are going to help improve urban education.

ROBERT STAKE,
DIRECTOR, THE CENTER FOR INSTRUCTIONAL RESEARCH AND CURRICULUM EVALUATION,
UNIVERSITY OF ILLINOIS, COLLEGE OF EDUCATION.

I will try to limit myself to three comments. We have been on quite a tour of urban education evaluation problems this morning and the variety is almost overwhelming. I will try to draw attention back to three things that our first three speakers said.

First was Norm Gold. How refreshing, to me, anyway, to hear an appeal to conceptualizing the program evaluation responsibility around the expectation of stakeholders. I say this not because I am, but that is not why. It is It is because that is the richest source of ideans for understanding what is going on in that program. I believe that Norm did say that the evaluations should be concerned about modifying the expectations, making them more realistic. I will argue with him on that one. I think we are too interventionist already in our evaluations and it is not our role to make expectations more realistic.

Next I will comment on Asa Hilliard's presentation. How delighted I was by much that he said about the use of observations, about the cultural deprivation of evaluators, and about content validity. I was distressed by his optimism that scholarly or measurement technology has in the foreseeable future too much of anything to contribute to program evaluation and by his failure to doubt more generally the common concept that aggregating information about individuals, their growth, their education, tells us much abut whether or not the program is operating in a responsible way according to ordinary expectations of what teachers and other people believe that should be doing.

The principal eye, it seems to me, of program evaluation should be on the professional educators, not on the students, asking partly whether those professional educators are attentive of those individual children.

Next I would comment on Eleanor Farrar's remarks. How delicious I thought it was to lead us up to the doors, telling us there is a tiger behind one and some more clues as to how to choose the doors. The implication was, of course, that if we in program evaluation see the evolving program as an evolving program, much more than as something that is fitted out at the factory, in Detroit or elsewhere, we will attend to the important issues, considerations, valuings and language -- and much of it will be street language -- that is there, and that this is our job. I see this to be the case and there are various types of case studies, ethnographic studies, and other studies, as has been implied elsewhere this morning, that are most useful in making this record.

There still is a grand expectation that I don't know what we can deal with. It was raised in the discussion first off this morning. There is an expectation that the program evaluation is going to uncover some secret of innovative education, some secret of teaching, some understanding about the processes that has not been recognized before, the aim of most research studies.

Program evaluators are not going to discover anything more with the time and budget constraints they have. Perhaps even if they did not have those constraints, they are not going to discover more than is common knowledge to different persons involved with that program. The secrets have to be looked for. We have to try to understand teaching and all that. But that is not going to happen within program evaluation, and it should not be organized around discovering those secrets.

What we can as program evaluators do is capture some of the experiences, the insights, the judgments of many people, right there at the site, and of some who are looking on more distantly, and make a sensible composite

of these with some observations and scores of our own and we will use extracts from documents we can pick up to tell a story of that particular program, its context, its problems, its pride and joy, and to leave largely for others to make the more final decisions as to the success and failure and desirability of use elsewhere of that program.

DISCUSSION

GWENDOLYN AUSTIN, Teacher Corps: I wanted to direct my comments to Ms. Lamoureaux. I think you have done an excellent job. I think that you are an asset to the school. But there is one request that I would like to make. I really became concerned when you made the statement about the pregnant student that her opportunity is over and that there is no hope. That hurt me because if this kid in the eight grade is lost, then there are a whole lot of them who are lost throughout the nation. There are a whole lot like her. These kids, with the encouragement of their parents, many of them have been able to pick themselves up and go on to greater things.

So, if PUSH is doing such a good job with parental involvement, then it needs to extend itself outside the school, to get to that parent and encourage that parent and apprise that parent of the opportunities that are out there to help that kid, raise her hopes and expectations, and development potential.

KAREN LAMOUREAUX: Thank you for your compliments. This is why I said that I hope sincerely that PUSH will be allowed to continue because somewhere along the line this little girl's failure was due to the lack of support at home. Somewhere along the line maybe PUSH can get to the parent.

I didn't mean that a pregnancy causes all students to fail. I was thinking really particularly of this one little girl. Given her character, the lack of support at home, and the pregnancy, her individual chances for succeeding are not assured. So, please, don't think that I believe that anybody must fail. But her case really got to me and I guess I was a little emotional about it.

YOLANDA BYARS: I want to say that PUSH/EXCEL has gotten into contact with Tammy's mother and it still hasn't done what it should have done. PUSH/EXCEL is a great program at Alton Park, but to a certain extent, as in Tammy's case, it was nothing.

JIM STEPHENSON, Teacher Corps: On your concerns about the role of the stakeholder and given the fact that there is no indication that this agency, the federal level, has in the past or would see nothing pending for the future. Would you comment, Dr. Stake on that problem, versus the continuation of a national large-scale impact evaluation?

DR. STAKE: In the past, there has been a great emphasis on the conceptualizations, the terms, the criteria of sucess and the like in terms of program designers and developers. Some people say well, those are important stakeholders, and, to be sure, they are. Other stakeholders at the scene, those people who are going to consider the problem as possibly useful in their school next year, the parents of the present children or the parents of children in the future, those stakeholders largely have been unattended to in the program evaluation efforts. I believe it was Norm's contention, with which I certainly agree, that, although it complicates things terribly, it is a reality with which we have failed to deal and with which we should be dealing more, the multiple expectations of different stakeholder groups.

SHIRLEY JACKSON, Director, Basic Skills Program: I guess I hear a bit of confusion, and I think this is the state of the art, in terms of doing naturalistic research, ethnographic research, and the preference there and the need for identifying student achievement gains. I am hearing a dichotomy there. We are saying both. In the school effectiveness studies, one of the things that you get all the time is the assessment and the monitoring of student progress. So, somehow or other we are going to have to intellectually put those two pieces together, and I'm not sure that we have yet.

DR. COMER: The monitoring of school achievement and performance I think is important and determining the criteria of what determines success is important. What I keep worrying about and what I did not hear even today is the question of what is important in moving from all of the things that we know to be involved in successful programs. What is required to remove unsuccessful programs from where they are to where they need to be? That, it seems to me, is something to which we are not paying enough attention. It is involved in the dichotomy about which you are concerned.

It seems to me, that we have observed that are necessary and critical and must be involved and which we must pay attention to if we are going to learn how to move unsuccessful programs to successful programs. There are specific interventions that are important in doing that.

DR. STAKE: I was more optimistic four years ago that we could work out technical procedures whereby the more scalar psychometric reading score records could be combined with observations and that we could validate one against the other. I am much less optimistic now. We have tried to do it. We may not be smart enough in my group. Others should keep trying. But we are not able to put them together very well at the present time.

My feeling is that it is better at the moment to lean on something that I believe is deliverable, that is, good representations of the process with the attendant information about individual case, classroom, school achievement; but that you are not going to be effectively able to aggregate that across schools, cities, and the country to get a picture of success of any urban program or any other educational program. The aggregation skills are not with us at the present time.

Both of the curiosities for the process and the impact are legitimate and should be pursued. Our ability to deliver at the present time I believe is much more in the area of describing what is going on in the teaching and activity of the classroom and the school and much less in the area of the reading scores.